

Approximation mit rationalen B-Spline Kurven und Flächen

Vom Fachbereich Mathematik
der Technischen Universität Darmstadt
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)
genehmigte

Dissertation

von
Dipl.-Math. Bernhard Elsässer
aus Würzburg

Referent:	Prof. Dr. J. Hoschek
Korreferent:	Prof. Dr. E. Hartmann
Tag der Einreichung:	23. 4. 1998
Tag der mündlichen Prüfung	26. 6. 1998

Darmstadt 1998
D 17

Inhaltsverzeichnis

Einleitung	1
1 Approximation mit rationalen B-Splinekurven	3
1.1 Das Approximationsproblem	4
1.2 Bestimmung einer Zielfunktion	5
1.3 Lösen des Approximationsproblems	14
1.3.1 Existenz und Eindeutigkeit der Lösung	14
1.3.2 Bestimmung einer Lösung	15
1.3.3 Parameterkorrektur	19
1.4 Beispiele	20
2 Approximation mit rationalen B-Splineflächen	29
2.1 Das Approximationsproblem	29
2.2 Lösungsverfahren	30
2.3 Beispiele	30
2.3.1 Rat. Approximationen im Vergleich	30
2.3.2 Test von rat. und int. Approximation	36
3 Segmentbestimmung	45
3.1 Iterative Segmentbestimmung	46
3.2 Ergebnisse	49
4 Approximation von Flächen mit Trimmkurven	61
4.1 neue Zielfunktion	62
4.2 Testresultate	64
4.2.1 Approximationsverhalten	64
4.2.2 Verfahren im Vergleich	73
Schlußwort	83
Literaturverzeichnis	85

Einleitung

Die vorliegende Arbeit behandelt Fragestellungen, die zum Themenkreis der Approximation von gegebenen Punktdaten mit rationalen B-Splinekurven oder -flächen gehören. Rationale Darstellungen von Freiformkurven und -flächen werden zunehmend in CAD-Systemen eingesetzt, da sie verglichen mit integralen Kurven und Flächen bei gleicher Segmentanzahl mehr Freiheitsgrade besitzen, die zum Beispiel als zusätzliche Designparameter genutzt werden können, und Kegelschnitte, Quadriken sowie Rotationsflächen exakt darstellen können (s. u. a. [19]). In einem Teilgebiet des CAGD (Computer Aided Geometric Design), dem Reverse Engineering, in dem ausgehend von mit einem Meßsystem abgetasteten Punkten der Oberfläche eines Objekts, z. B. eines Karosserieteils eines Autos, versucht wird diese Oberfläche als mathematische Flächendarstellung zu rekonstruieren, spielt daher die Approximation von Punktdaten eine zentrale Rolle. Auch hier können rationale B-Splinedarstellungen aufgrund ihrer höheren Anzahl von Freiheitsgraden vorteilhaft verwendet werden, z. B. um rotationssymmetrisch gelegene Punkte mit exakten Rotationsflächen zu approximieren ([6], [20]).

Bisher bekannte Verfahren (z. B. [14], [16], [22], [25]), die rationale B-Splines zur Approximation verwenden, lassen sich wie folgt charakterisieren: [14], [25] (s. a. Kapitel 1) benutzen numerische Methoden (Newton-, Gradientenverfahren), um die Fehlerquadratsumme $\sum(\mathbf{X}() - \mathbf{P}_i)^2$ zu minimieren (nichtlineares Optimierungsproblem). $\mathbf{X}() := \mathbf{R}()/S()$ ist die Parameterdarstellung einer rationalen B-Splinekurve oder -fläche und \mathbf{P}_i sind die vorgegebenen Punkte. Der Zähler $\mathbf{R}()$ und der Nenner $S()$ bestehen aus Linearkombinationen von B-Splinefunktionen, deren vektorwertige (de Boor-Punkte) oder skalare Koeffizienten (Gewichte) die Optimierungsvariablen sind. [16] geht von den linearen Gleichungen $\mathbf{R}() = S() \mathbf{P}_i$ aus, um die de Boor-Punkte und Gewichte zu bestimmen. Bei der Lösung dieses (überbestimmten) Gleichungssystems müssen mehrere lineare Systeme, aber auch ein nichtlineares Optimierungsproblem gelöst werden, um positive Gewichte zu gewährleisten. Die Bestimmung positiver Gewichte bei der Approximation mit rationalen B-Splines ist eine für die Praxis unbedingt notwendige Forderung (s. u.). In [22] (s. Kapitel 1) entsteht durch Multiplikation mit dem Nenner $S()$ ein lineares Optimierungsproblem. Alle eben beschriebenen Verfahren besitzen einen oder mehrere der folgenden Nachteile: hoher Berechnungsaufwand, hoher Speicherplatzbedarf, numerische Verfahrensfehler, Beschränkung der Ordnung der B-Splinefunktionen und unbrauchbare Lösungen (z. B. Nullstellen im Nenner $S()$). Wegen dieser Nachteile ist das jetzt zu präsentierende rationale Approximationsverfahren mit dem Ziel entwickelt worden, daß es einen möglichst geringen Rechenaufwand (lineares Optimierungsproblem) hat, einen geringen Speicherverbrauch hat, die Ordnung der B-Splinefunktionen keiner Beschränkung unterliegt und es positive Gewichte liefert. Dabei sollen die de Boor-

Punkte und die Gewichte Optimierungsvariablen sein. Dieses Verfahren hält die Zielvorgaben ein. In den folgenden Kapiteln wird es nun zusammen mit daraus abgeleiteten oder darauf aufbauenden Verfahren, die zur Approximation mit rationalen B-Splinekurven und -flächen mit Erfolg einsetzbar sind, im Detail vorgestellt.

Das erste Kapitel befaßt sich mit der Approximation mit rationalen B-Splinekurven. Hier wird mittels homogener Koordinaten eine Zielfunktion entwickelt, deren Minimierung auf ein reguläres lineares System führt und positive Gewichte als Lösungen sicherstellt. Die Garantie positiver Gewichte bei rationalen B-Splinekurven hat zur Folge, daß sie für die Praxis wichtige geometrische Eigenschaften wie z. B. konvexe Hülle-Eigenschaft usw. (s. [11]) der integralen Kurven beibehalten. Weiter wird gezeigt, daß diese Zielfunktion für Gewichte gleich Eins näherungsweise euklidische Abstandsquadrate beschreibt.

Im nächsten Kapitel werden die Ergebnisse von Kapitel 1 auf die Approximation mit rationalen B-Splineflächen übertragen, um auch ein Approximationsverfahren für den wichtigen Flächenfall zu erhalten.

In Kapitel 3 wird ein Lösungsansatz für das Problem der Bestimmung der Segmentanzahl für die B-Splineapproximation mit Kurven (rational oder integral) erläutert. Bei vorgegebener Fehlertoleranz ε wird mit einem abgewandelten Sekantenverfahren eine Stelle \bar{s} der Funktion $f(s) := F_s - \varepsilon$ mit $f(\bar{s}) \leq 0$ bestimmt, wobei F_s die Summe der Fehlerquadrate oder der maximale Fehler zwischen der Approximationskurve mit s Segmenten und den Vorgabepunkten ist. Somit ist eine \bar{s} -segmentige Approximationskurve gefunden, die die Fehlertoleranz ε einhält.

Das Schlußkapitel behandelt die Approximation von vorgegebenen Flächenpunkten $\mathbf{P}_i = \mathbf{Y}(u_i, v_i)$ mit einer Trimmkurve $\mathbf{Y}(\mathbf{C}(t))$. $\mathbf{Y}(u, v)$ ist die Parameterdarstellung einer Fläche und $\mathbf{C}(t)$ die Parameterdarstellung einer rationalen B-Splinekurve in der Parameterebene der Fläche. Durch Linearisierung des Fehlerquadrats $(\mathbf{Y}(\mathbf{C}(t_i)) - \mathbf{P}_i)^2$ und geeignete Abschätzung lassen sich die Ergebnisse aus Kapitel 1 übertragen, so daß ein lineares Optimierungsproblem entsteht.

Mein besonderer Dank gilt Herrn Prof. Dr. J. Hoschek für die freundliche Unterstützung sowie die zahlreichen Anregungen und Hinweise, die diese Arbeit mit ermöglicht haben. Weiter möchte ich mich bei den Sekretärinnen Frau E. Kniffki und Frau S. Drexler und allen Mitarbeitern der AG 3, u. a. bei Herrn Dipl.-Math. U. Dietz, der mir die Beispiele der integralen Flächenapproximation in Abschnitt 2.3.2 mit seinem Verfahren erstellt hat, für die angenehme und hilfsbereite Atmosphäre bedanken.

Kapitel 1

Approximation mit rationalen B-Splinekurven

Zum Thema Approximation mit rationalen Kurven sind im Rahmen des CAGD schon einige Beiträge veröffentlicht worden ([14], [16], [22], [25]). Bei dieser Fragestellung handelt es sich kurz gesagt darum, eine rationale Kurve $\mathbf{X}(t) := \mathbf{R}(t)/S(t) \in \mathbb{R}^l$ ($l = 2, 3$) zu finden, die gegebene Punkte $\mathbf{P}_i \in \mathbb{R}^l$ ($i = 0, \dots, m$) im diskreten Fall oder eine gegebene Kurve $\mathbf{f}(t) \in \mathbb{R}^l$ im kontinuierlichen Fall möglichst gut approximiert. Der Zähler $\mathbf{R}(t)$ und der Nenner $S(t)$ bestehen aus polynomialen Basis(-Spline)funktionen, deren vektorwertige oder skalare Koeffizienten bei der Approximation zu bestimmen sind.

Diesen Themenbereich behandeln unter anderem die Arbeiten von [25] und [22]. In [25] werden die euklidischen Abstandsquadrate

$$\sum_{i=0}^m (\mathbf{X}(t_i) - \mathbf{P}_i)^2$$

minimiert. Dieses nichtlineare Optimierungsproblem wird in aufwendigen Schritten mit Hilfe der Pseudo-Inversen einer Matrix linearisiert und dann iterativ gelöst (s. [25]). Das Verfahren hat folgende gewichtige Nachteile: einen hohen Rechenaufwand, der sich durch lange CPU-Zeiten insbesondere bei großen Punktzahlen ($i > 100$) bemerkbar macht, hoher Speicherplatzverbrauch bei großen Punktzahlen und die Beschränkung auf Polynomgrade < 4 .

[22] verfolgt einen anderen Ansatz. Um von vornherein ein nichtlineares Problem zu vermeiden, wird mit dem Nenner $S(t)$ von $\mathbf{X}(t)$ durchmultipliziert und im kontinuierlichen Fall

$$\int_a^b (\mathbf{R}(t) - S(t) \mathbf{f}(t))^2 dt$$

minimiert. Der Ansatz ist leicht auf den diskreten Fall übertragbar. Das Verfahren hat einen geringen Rechenaufwand (Lösung eines linearen Gleichungssystems). Negativ ins Gewicht fällt aber die Tatsache, daß das Verfahren in manchen Fällen dazu neigt, den Nenner $S(t)$ zu Null werden zu lassen. Das hat für die rationale Kurve Polstellen zur Folge, so daß die Kurve für Anwendungen unbrauchbar ist.

Aufgrund der dargestellten Probleme bei diesen Methoden wurde das nachstehende Approximationsverfahren unter den folgenden Vorgaben entwickelt:

- (i) de Boor-Punkte und Gewichte sind Optimierungsparameter
- (ii) möglichst geringer Rechenaufwand (lineares Optimierungsproblem)
- (iii) Ordnung der B-Splinekurve unterliegt (prinzipiell) keiner Beschränkung
- (iv) es sollen positive Gewichte erzeugt werden, um die bekannten Eigenschaften der B-Splines zu erfüllen

1.1 Das Approximationsproblem

Gegeben sind $m+1$ Punkte im reellen euklidischen Raum E^3 mit den kartesischen Koordinaten $\mathbf{P}_i := (p_{ix}, p_{iy}, p_{iz})^\top \in \mathbb{R}^3$ und den Parameterwerten t_i , für die gilt $t_0 < t_1 < \dots < t_m$. Gesucht ist die rationale B-Splinekurve der Ordnung k (s. a. [11]) gegeben durch

$$\mathbf{X}(t) := \frac{\sum_{j=0}^n h_j \mathbf{d}_j N_{jk}(t)}{\sum_{j=0}^n h_j N_{jk}(t)} \quad , \quad \mathbf{d}_j := \begin{pmatrix} d_{jx} \\ d_{jy} \\ d_{jz} \end{pmatrix}$$

Die \mathbf{d}_j sind die de Boor-Punkte und die h_j die Gewichte. Die B-Splinefunktionen N_{jk} der Ordnung k sind dabei über dem Knotenvektor

$$\mathbf{T} := (\tau_0 = \dots = \tau_{k-1}, \tau_k, \dots, \tau_n, \tau_{n+1} = \dots = \tau_{n+k}) \quad , \quad n \geq k-1$$

definiert und es gilt $n < m$ sowie $t_0 = \tau_0, t_m = \tau_{n+1}$. Außerdem sollen die Parameterwerte t_i die *Schönberg-Whitney* Bedingungen erfüllen, die später in Abschnitt 1.2 benötigt werden: es muß also eine geordnete Teilmenge $\{s_0, \dots, s_n\} \subset \{t_0, \dots, t_m\}$ mit $s_0 < s_1 < \dots < s_n$ existieren, so daß

$$\tau_j < s_j < \tau_{j+k} \quad (j = 0, \dots, n)$$

gilt. Nun können Punkte des E^3 durch Verwendung *homogener Koordinaten* in den reellen projektiven Raum P^3 eingebettet werden: Ein Punkt im E^3 mit den kartesischen Koordinaten

$$\mathbf{Y} = \begin{pmatrix} y_x \\ y_y \\ y_z \end{pmatrix}$$

erhält dann im P^3 die homogenen Koordinaten

$$\bar{\mathbf{Y}} = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ y_x \\ y_y \\ y_z \end{pmatrix} \quad (\alpha \in \mathbb{R} \setminus \{0\})$$

Das Approximationsproblem besteht jetzt darin, eine geeignete Zielfunktion

$$\bar{F} := F + \lambda G \longrightarrow \min \quad (\bar{F} \in \mathbb{C}^2, \quad \lambda > 0, \quad \text{fest gewählt}) \quad (1.1)$$

mit

$$F := \sum_{i=0}^m Q(\bar{\mathbf{X}}(t_i), \bar{\mathbf{P}}_i)$$

und der Nebenbedingung G zu minimieren, so daß die rationale B-Splinekurve die gegebenen Punkte möglichst gut approximiert. In (1.1) sind $\bar{\mathbf{X}}(t)$ und $\bar{\mathbf{P}}_i$ homogene Koordinaten der gesuchten rationalen B-Splinekurve und der gegebenen Punkte. Die Funktion $Q \geq 0$ dient dazu, einen Abstand zwischen zwei Punkten in homogenen Koordinaten beschreiben, G soll ein Gewichtskontrollterm sein und der Faktor λ steuert den Einfluß des Kontrollterms. Die Einführung von G und λ hat zum Ziel, daß die Minimierung von (1.1) positive Gewichte h_j liefert. Im nächsten Abschnitt wird die Entwicklung einer geeigneten Zielfunktion dargestellt.

1.2 Bestimmung einer Zielfunktion

In (1.1) wird gewählt

$$\bar{\mathbf{X}}(t) := \begin{pmatrix} \sum_{j=0}^n h_j N_{jk}(t) \\ \sum_{j=0}^n \bar{d}_{jx} N_{jk}(t) \\ \sum_{j=0}^n \bar{d}_{jy} N_{jk}(t) \\ \sum_{j=0}^n \bar{d}_{jz} N_{jk}(t) \end{pmatrix}, \quad \bar{\mathbf{P}}_i := c_i \begin{pmatrix} 1 \\ p_{ix} \\ p_{iy} \\ p_{iz} \end{pmatrix} \quad (1.2)$$

mit $\bar{d}_{jx} := h_j d_{jx}$, $\bar{d}_{jy} := h_j d_{jy}$, $\bar{d}_{jz} := h_j d_{jz}$ und noch festzulegenden c_i .

In der Einführung des Kapitels wurde unter anderem gefordert, ein lineares Optimierungsproblem zu entwerfen und damit einen günstigen Rechenaufwand zu haben. Um beim Aufstellen der notwendigen Bedingungen zur Minimierung von (1.1) ein lineares System zu erhalten, dürfen die Optimierungsvariablen h_j , \bar{d}_{jx} , \bar{d}_{jy} und \bar{d}_{jz} nur linear oder quadratisch in Q auftreten. Funktionen dieser Art sind im wesentlichen quadratische Formen in $\bar{\mathbf{X}}(t)$. Für Q wird daher angesetzt

$$Q(\bar{\mathbf{X}}, \bar{\mathbf{Y}}) := \sum_{i,j,k,l=0}^3 a_{ij}^{kl} x_i x_j y_k y_l \quad (1.3)$$

In (1.3) sind $\bar{\mathbf{X}}, \bar{\mathbf{Y}}$ homogene Koordinaten zweier beliebiger Punkte in P^3 .

Wir stellen folgende sinnvolle Forderungen an Q :

(i) Q ist eine quadratische Form in $\bar{\mathbf{X}}$ bzw. in $\bar{\mathbf{Y}}$.

$$\implies a_{ij}^{kl} = a_{ji}^{kl}, \quad a_{ij}^{kl} = a_{ij}^{lk}$$

(ii) $Q(\bar{\mathbf{X}}, \bar{\mathbf{Y}}) = Q(\bar{\mathbf{Y}}, \bar{\mathbf{X}})$ (Symmetrie)

$$\implies a_{ij}^{kl} = a_{kl}^{ij}$$

(iii) Beschreiben die Koordinaten $\bar{\mathbf{X}}$ und $\bar{\mathbf{Y}}$ den selben Punkt, so soll Q verschwinden, d. h.

$$Q(\bar{\mathbf{X}}, \alpha \bar{\mathbf{X}}) = \alpha^2 \sum_{i,j,k,l=0}^3 a_{ij}^{kl} x_i x_j x_k x_l = 0 \quad (\text{Interpolationseigenschaft})$$

Aus der letzten Forderung ergeben sich folgende Gleichungen für die Koeffizienten a_{ij}^{kl} :

$$8a_{ij}^{kl} + 8a_{ik}^{jl} + 8a_{il}^{jk} = 0 \quad (k \neq l \neq i \neq j) \quad (1.4)$$

$$4a_{ij}^{kk} + 8a_{ik}^{kj} = 0 \quad (k \neq i \neq j) \quad (1.5)$$

$$2a_{ii}^{kk} + 4a_{ik}^{ik} = 0 \quad (k \neq i) \quad (1.6)$$

$$4a_{ik}^{kk} = 0 \quad (k \neq i) \quad (1.7)$$

$$a_{kk}^{kk} = 0 \quad (1.8)$$

Man sieht das beispielsweise für die Gleichungen (1.5) so: Wegen (iii) müssen die Koeffizienten von $x_k^2 x_i x_j$ ($k \neq i \neq j$) aufsummiert Null sein. Aufgrund der Forderungen (i), (ii) gibt es nur die zwei verschiedenen Koeffizienten a_{ij}^{kk} , a_{ik}^{kj} . Der Koeffizient a_{ij}^{kk} tritt 4-fach, der Koeffizient a_{ik}^{kj} 8-fach auf. Deshalb gilt (1.5). Für $k \neq i \neq j$ hat man 12 Gleichungen und es gibt insgesamt $4 \cdot 12 + 8 \cdot 12 = 144$ Koeffizienten für $x_k^2 x_i x_j$. Auf diese Weise ergeben sich auch die restlichen Gleichungen (1.4),(1.6)-(1.8). In der folgenden Tabelle sind die Ergebnisse aus Forderung (iii) zusammengefasst. In der ersten Spalte sind die zu den Koeffizienten gehörigen Koordinatenprodukte, z. B. in der ersten Zeile $x_k x_l x_i x_j$, und der entsprechende Gleichungstyp aufgelistet. In der zweiten Spalte sieht man die Anzahl der Koeffizienten zum jeweiligen Koordinatenprodukt. Die dritte Spalte zeigt die Anzahl der verschiedenen Koeffizienten. In der letzten Spalte steht die Anzahl der Gleichungen zum Gleichungstyp aus der ersten Spalte. Die letzte Zeile der Tabelle zeigt in den Spalten 2,3 und 4 die jeweiligen aufsummierten Anzahlen. Aus der Differenz der Werte der dritten und vierten Spalte $55 - 35 = 20$ erhält man 20 Freiheitsgrade für die Wahl der Koeffizienten a_{ij}^{kl} .

Koeffizienten von, Gleichung	Anzahl der Koeffizienten	Anzahl der verschiedenen Koeffizienten	Anzahl der Gleichungen
$x_k x_l x_i x_j \quad (k \neq l \neq i \neq j)$ Gl. (1.4)	24	3	1
$x_k^2 x_i x_j \quad (k \neq i \neq j)$ Gl. (1.5)	144	24	12
$x_k^2 x_i^2 \quad (k \neq i)$ Gl. (1.6)	36	12	6
$x_k^3 x_i \quad (k \neq i)$ Gl. (1.7)	48	12	12
x_k^4 Gl. (1.8)	4	4	4
	$\Sigma = 256$	$\Sigma = 55$	$\Sigma = 35$

Werden die Koordinaten

$$\bar{\mathbf{z}} = \begin{pmatrix} z_0 \\ z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \end{pmatrix} = \begin{pmatrix} x_0 y_1 - x_1 y_0 \\ x_0 y_2 - x_2 y_0 \\ x_0 y_3 - x_3 y_0 \\ x_1 y_2 - x_2 y_1 \\ x_1 y_3 - x_3 y_1 \\ x_2 y_3 - x_3 y_2 \end{pmatrix} \quad (1.9)$$

in Verbindung mit der quadratischen Form Q^* des \mathbb{R}^6

$$Q^*(\bar{\mathbf{z}}) = \bar{\mathbf{z}}^T \mathbf{C} \bar{\mathbf{z}} = \sum_{i,j=0}^5 c_{ij} z_i z_j, \quad \mathbf{C} = \mathbf{C}^T = (c_{ij}) \quad (i, j = 0, \dots, 5) \quad (1.10)$$

eingeführt, so zeigt sich durch Umrechnen in die Darstellung (1.3), daß $Q^*(\bar{\mathbf{z}})$ die Gleichungen (1.4)-(1.8) erfüllt. Gibt man sich umgekehrt in den Gleichungen (1.4) z. B. a_{01}^{23} und a_{02}^{13} , in (1.5) die a_{ij}^{kk} und in (1.6) die a_{ii}^{kk} vor (20 Freiheitsgrade), dann erhält man

$$\begin{pmatrix} c_{05} \\ c_{14} \\ c_{23} \end{pmatrix} = \begin{pmatrix} 2a_{02}^{13} \\ 2a_{01}^{23} \\ 0 \end{pmatrix} + a \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \quad (a \in \mathbb{R}) \quad (1.11)$$

$$\begin{array}{lll} c_{01} = a_{12}^{00} & c_{02} = a_{13}^{00} & c_{12} = a_{23}^{00} \\ c_{03} = -a_{02}^{11} & c_{04} = -a_{03}^{11} & c_{34} = a_{23}^{11} \\ c_{13} = a_{01}^{22} & c_{15} = -a_{03}^{22} & c_{35} = -a_{13}^{22} \\ c_{24} = a_{01}^{33} & c_{25} = a_{02}^{33} & c_{45} = a_{12}^{33} \\ c_{00} = a_{11}^{00} & c_{11} = a_{22}^{00} & c_{22} = a_{33}^{00} \\ c_{33} = a_{22}^{11} & c_{44} = a_{33}^{11} & c_{55} = a_{33}^{22} \end{array}$$

Die bisherigen Überlegungen lassen sich zusammenfassen in

Satz 1.2.1 *Jede Funktion Q der Form (1.3), die die Bedingungen (i)–(iii) erfüllt, läßt sich durch eine quadratische Form $Q^*(\bar{\mathbf{z}}) = \bar{\mathbf{z}}^T \mathbf{C} \bar{\mathbf{z}}$ mit $\bar{\mathbf{z}}$ aus (1.9) darstellen.*

Setzt man in (1.11) $a_{02}^{13} = a_{01}^{23} = 0$ und $a = 1$ folgt

$$z_0 z_5 - z_1 z_4 + z_2 z_3 = 0 \quad (1.12)$$

In der Literatur, z. B. [10], werden die Koordinaten (1.9) als *Plücker-Koordinaten* und die Gleichung (1.12) als *Plücker-Quadrik* bezeichnet (s. a. [21]). Die Plücker-Koordinaten können als homogene Koordinaten eines Punkts des 5-dimensionalen reellen projektiven Raums P^5 gedeutet werden. Genauer gesagt: Ist g eine Gerade des P^3 , die durch zwei auf ihr liegenden Punkten mit den projektiven Koordinaten $\bar{\mathbf{X}}$ und $\bar{\mathbf{Y}}$ bestimmt ist, dann wird g , gegeben durch $(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$, mit $f : (\bar{\mathbf{X}}, \bar{\mathbf{Y}}) \mapsto \bar{\mathbf{z}}$ auf einen Punkt $G \in P^5$ mit den Plücker-Koordinaten $\bar{\mathbf{z}}$ abgebildet. G liegt auf der Plücker-Quadrik. Die Abbildung f ist sinnvoll definiert, d. h. unabhängig von der Wahl der beiden Geradenpunkte und Multiplikation der Punktkoordinaten mit einer reellen Zahl ungleich Null ist das Bild von g unter f immer

derselbe Punkt G . Die Geraden des P^3 werden durch f bijektiv auf die Punkte der Plücker-Quadrik abgebildet.

Im Hinblick auf das Minimierungsproblem (1.1) wird von Q in (1.3) noch verlangt

$$(iv) \quad Q(\bar{\mathbf{X}}, \bar{\mathbf{Y}}) > 0, \text{ wenn } \bar{\mathbf{Y}} \neq \alpha \bar{\mathbf{X}} \quad (\alpha \neq 0, \quad \bar{\mathbf{X}}, \bar{\mathbf{Y}} \neq \mathbf{0})$$

Da für die Plücker-Koordinaten (1.9) gilt:

$$\bar{\mathbf{Y}} = \alpha \bar{\mathbf{X}} \implies \bar{\mathbf{z}} = \mathbf{0},$$

folgt, daß die Matrix \mathbf{C} in (1.10) vollen Rang hat und der Bedingung $\bar{\mathbf{z}}^T \mathbf{C} \bar{\mathbf{z}} > 0$ für $\bar{\mathbf{z}} \neq \mathbf{0}$ genügt. Damit erhalten wir das

Korollar 1.2.2 *Erfüllt Q die Bedingungen (i)–(iv), dann ist Q^* eine positiv definite quadratische Form.*

Es genügt aufgrund des später folgenden Satzes 1.2.3, die Matrix \mathbf{C} als Einheitsmatrix, der einfachsten positiv definiten Matrix, zu wählen. Die Abstandsfunktion F besitzt dann die Gestalt

$$F = \sum_{i=0}^m \bar{\mathbf{P}}_i^2 \left(\bar{\mathbf{X}}^2(t_i) - \frac{(\bar{\mathbf{X}}(t_i) \cdot \bar{\mathbf{P}}_i)^2}{\bar{\mathbf{P}}_i^2} \right)$$

mit \cdot als Skalarprodukt des \mathbb{R}^4 . Damit F unabhängig vom Betragsquadrat der $\bar{\mathbf{P}}_i$ wird, werden die c_i aus (1.2) als Normierungsfaktoren gewählt. Die Approximationspunkte mögen also die homogenen Koordinaten

$$\bar{\mathbf{P}}_i = \frac{1}{\sqrt{1 + p_{ix}^2 + p_{iy}^2 + p_{iz}^2}} \begin{pmatrix} 1 \\ p_{ix} \\ p_{iy} \\ p_{iz} \end{pmatrix} \quad (1.13)$$

haben und mit den Bezeichnungen

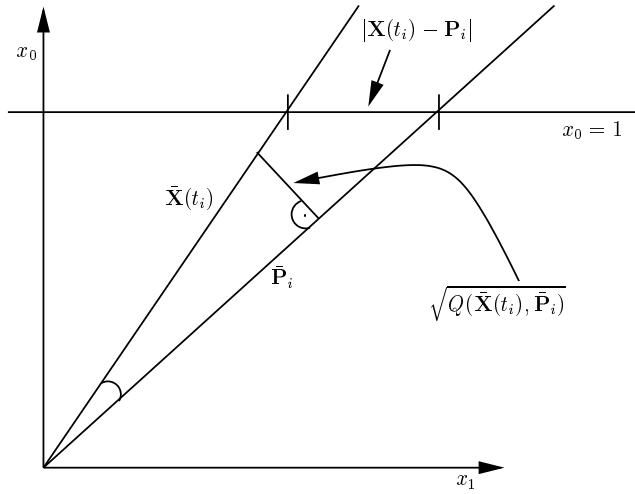
$$\bar{\mathbf{X}}(t) = \begin{pmatrix} x_0(t) \\ \bar{\mathbf{x}}(t) \end{pmatrix}, \quad \bar{\mathbf{x}}(t) = \begin{pmatrix} \sum_{j=0}^n \bar{d}_{jx} N_{jk}(t) \\ \sum_{j=0}^n \bar{d}_{jy} N_{jk}(t) \\ \sum_{j=0}^n \bar{d}_{jz} N_{jk}(t) \end{pmatrix}, \quad x_0(t) = \sum_{j=0}^n h_j N_{jk}(t)$$

hat die Abstandsfunktion F schließlich die Form

$$\begin{aligned} F &= \sum_{i=0}^m \bar{\mathbf{X}}^2(t_i) - (\bar{\mathbf{X}}(t_i) \cdot \bar{\mathbf{P}}_i)^2 \\ &= \sum_{i=0}^m x_0^2(t_i) + \bar{\mathbf{x}}^2(t_i) - \frac{1}{1 + \bar{\mathbf{P}}_i^2} (x_0(t_i) + \bar{\mathbf{x}}(t_i) \cdot \bar{\mathbf{P}}_i)^2 \end{aligned} \quad (1.14)$$

F bzw. Q kann geometrisch gedeutet werden: Abbildung 1.1 zeigt die Situation im P^3 in der x_0 - x_1 -Ebene. Die Punkte $\bar{\mathbf{X}}(t_i)$ und $\bar{\mathbf{P}}_i$ sind Geraden durch den Ursprung. Der Abstand Q kann hier so formuliert werden

$$Q(\bar{\mathbf{X}}(t_i), \bar{\mathbf{P}}_i) = \bar{\mathbf{X}}^2(t_i) - (\bar{\mathbf{X}}(t_i) \cdot \bar{\mathbf{P}}_i)^2 = \sin^2(\angle \bar{\mathbf{X}}(t_i), \bar{\mathbf{P}}_i) \bar{\mathbf{X}}^2(t_i)$$

Abbildung 1.1: Geometrische Deutung von F bzw. Q

und \sqrt{Q} ist damit die Gegenkathete des eingezeichneten rechtwinkligen Dreiecks. Man erkennt, daß für vom Betrag her kleine x_1 -Koordinaten von $\tilde{\mathbf{X}}(t_i)$, $\bar{\mathbf{P}}_i$, d. h. $\mathbf{X}(t_i)$, \mathbf{P}_i befinden sich in der Nähe des Koordinatenursprungs, kleine Abstände \sqrt{Q} auch kleine euklidische Abstände $|\mathbf{X}(t_i) - \mathbf{P}_i|$ bedeuten. Diese wichtige Beobachtung wird unten im Satz 1.2.3 präzisiert.

F ist eine quadratische Form in den Variablen h_j , \tilde{d}_{jx} , \tilde{d}_{jy} und \tilde{d}_{jz} . Indem die Parameterwerte t_i den *Schönberg-Whitney* Bedingungen genügen, ist gesichert, daß aus $\tilde{\mathbf{X}}(t_i) = \mathbf{0}$ ($i = 0, \dots, m$) folgt $h_j = 0$, $\tilde{d}_{jx} = 0$, $\tilde{d}_{jy} = 0$ und $\tilde{d}_{jz} = 0$. Weil F den Bedingungen (iii), (iv) genügt, ist F positiv semidefinit. F ist sogar positiv definit, wenn die rationale B-Splinekurve mindestens einen der gegebenen Punkte $\bar{\mathbf{P}}_i$ ($i = 0, \dots, m$) nicht interpolieren kann.

F ist invariant gegenüber Drehungen des E^3 : Ist \mathbf{D} eine orthogonale Drehmatrix, dann ergibt sich

$$\begin{aligned} \sum_{i=0}^m x_0^2(t_i) + \bar{\mathbf{x}}^\top(t_i) \mathbf{D}^\top \mathbf{D} \bar{\mathbf{x}}(t_i) - \frac{1}{1 + \mathbf{P}_i^\top \mathbf{D}^\top \mathbf{D} \mathbf{P}_i} (x_0(t_i) + \bar{\mathbf{x}}^\top(t_i) \mathbf{D}^\top \mathbf{D} \mathbf{P}_i)^2 \\ = \sum_{i=0}^m x_0^2(t_i) + \bar{\mathbf{x}}^2(t_i) - \frac{1}{1 + \mathbf{P}_i^2} (x_0(t_i) + \bar{\mathbf{x}}(t_i) \cdot \mathbf{P}_i)^2 \end{aligned}$$

F ist aber nicht invariant gegenüber Translationen des E^3 . Es ist jedoch für die Approximation aus geometrischer Sicht eine sehr wesentliche Forderung, daß die rationale B-Splinekurve $\mathbf{X}(t)$ als Lösung des Minimierungsproblems (1.1) ähnlichkeitsinvariant zu den gegebenen

Punkten ist. Um dieser Forderung zu genügen, wird deshalb vor dem Approximationsprozeß die folgende Koordinatentransformation durchgeführt:

- (i) der Koordinatenursprung des neuen Koordinatensystems ist der Schwerpunkt der Punkte \mathbf{P}_i
- (ii) gleichmäßige Skalierung der Punktkoordinaten so, daß $|\mathbf{P}_i| \leq d \quad (i = 0, \dots, m)$

Die Konstante d wird später festgelegt. Nach dem Approximationsprozeß werden die Punkte \mathbf{P}_i und die gewonnene rationale B-Splinekurve mit der zu (i), (ii) inversen Koordinatentransformation zurücktransformiert.

Eine weitere wesentliche Eigenschaft, die die Wahl von F rechtfertigt, resultiert aus dem nächsten

Satz 1.2.3 *Im Fall einer integralen B-Splinekurve $\mathbf{X}(t) = \bar{\mathbf{x}}(t)$ und Approximationspunkten \mathbf{P}_i mit Norm $|\mathbf{P}_i| < \varepsilon \quad (i = 0, \dots, m)$ beschreibt F näherungsweise die Summe der euklidischen Abstandsquadrate zwischen den Kurven- und den Approximationspunkten. Die Konvergenzordnung von (1.16) ist $O(|\mathbf{P}_i|^2)$.*

Bevor wir den Satz beweisen, stellen wir den folgenden Hilfssatz bereit:

Hilfssatz 1.2.4 *Die Summanden von F erfüllen die Gleichung*

$$\bar{\mathbf{X}}^2(t_i) - (\bar{\mathbf{X}}(t_i) \cdot \bar{\mathbf{P}}_i)^2 = (\bar{\mathbf{x}}(t_i) - x_0(t_i)\mathbf{P}_i)^2 - \frac{\mathbf{P}_i^2}{1 + \mathbf{P}_i^2} \left((\bar{\mathbf{x}}(t_i) - x_0(t_i)\mathbf{P}_i) \cdot \frac{\mathbf{P}_i}{|\mathbf{P}_i|} \right)^2 \quad (1.15)$$

Beweis. Folgende Umformungen zeigen die Aussage des Hilfssatzes:

$$\begin{aligned} & \bar{\mathbf{X}}^2(t_i) - (\bar{\mathbf{X}}(t_i) \cdot \bar{\mathbf{P}}_i)^2 \\ &= x_0^2(t_i) + \bar{\mathbf{x}}^2(t_i) - \frac{1}{1 + \mathbf{P}_i^2} (x_0(t_i) + \bar{\mathbf{x}}(t_i) \cdot \mathbf{P}_i)^2 \\ &= x_0^2(t_i) + \bar{\mathbf{x}}^2(t_i) - \frac{1}{1 + \mathbf{P}_i^2} (x_0^2(t_i) + 2x_0(t_i)\bar{\mathbf{x}}(t_i) \cdot \mathbf{P}_i + (\bar{\mathbf{x}}(t_i) \cdot \mathbf{P}_i)^2) \\ &= \bar{\mathbf{x}}^2(t_i) - \frac{(\bar{\mathbf{x}}(t_i) \cdot \mathbf{P}_i)^2}{\mathbf{P}_i^2} + \frac{1}{1 + \mathbf{P}_i^2} \left(\frac{(\bar{\mathbf{x}}(t_i) \cdot \mathbf{P}_i)^2}{\mathbf{P}_i^2} - 2x_0(t_i)\bar{\mathbf{x}}(t_i) \cdot \mathbf{P}_i + x_0^2(t_i)\mathbf{P}_i^2 \right) \\ &= \left(\bar{\mathbf{x}}(t_i) \times \frac{\mathbf{P}_i}{|\mathbf{P}_i|} \right)^2 + \frac{1}{1 + \mathbf{P}_i^2} \left((\bar{\mathbf{x}}(t_i) - x_0(t_i)\mathbf{P}_i) \cdot \frac{\mathbf{P}_i}{|\mathbf{P}_i|} \right)^2 \\ &= \left((\bar{\mathbf{x}}(t_i) - x_0(t_i)\mathbf{P}_i) \times \frac{\mathbf{P}_i}{|\mathbf{P}_i|} \right)^2 + \frac{1}{1 + \mathbf{P}_i^2} \left((\bar{\mathbf{x}}(t_i) - x_0(t_i)\mathbf{P}_i) \cdot \frac{\mathbf{P}_i}{|\mathbf{P}_i|} \right)^2 \\ &= (\bar{\mathbf{x}}(t_i) - x_0(t_i)\mathbf{P}_i)^2 - \frac{\mathbf{P}_i^2}{1 + \mathbf{P}_i^2} \left((\bar{\mathbf{x}}(t_i) - x_0(t_i)\mathbf{P}_i) \cdot \frac{\mathbf{P}_i}{|\mathbf{P}_i|} \right)^2 \end{aligned}$$

$|\cdot|$ bezeichnet die euklidische Norm und \times das Vektorprodukt im \mathbb{R}^3 . ■

Mit der Aussage des Hilfssatzes können wir nun den Beweis des Satzes führen: Setzt man die

Gewichte $h_j = 1$ (integrale B-Splinekurve), und erfüllt ein Punkt \mathbf{P}_i die Bedingung $|\mathbf{P}_i| < \varepsilon$, so gilt

$$\bar{\mathbf{X}}^2(t_i) - (\bar{\mathbf{X}}(t_i) \cdot \bar{\mathbf{P}}_i)^2 \approx (\bar{\mathbf{x}}(t_i) - \mathbf{P}_i)^2 \quad (1.16)$$

Dies folgt aus Hilfssatz 1.2.4, wenn man in (1.15) den Term $-\mathbf{P}_i^2/(1+\mathbf{P}_i^2) \dots$ auf der rechten Seite vernachlässigt. Das ist jedoch nach obiger Voraussetzung gerechtfertigt. ■

Es genügt $\varepsilon = 1/2$ zu wählen, um (1.16) zu erfüllen. Eigenschaft (1.16) zeigt die Ähnlichkeit der nicht-euklidischen Abstandsfunktion F (1.14) zu den euklidischen Abstandsquadraten. Damit die Eigenschaft (1.16) zum Tragen kommt, wird im Skalierungsschritt (ii) der Koordinatentransformation $d = 1/2$ gewählt. Diese Skalierung stellt zusätzlich sicher, daß bei der Minimierung von (1.1) die Fehlervektoren (näherungsweise) senkrecht zur rationalen Approximationskurve stehen. Um die signifikanten Unterschiede zwischen der Approximation mit oder ohne Skalierung zu zeigen, betrachte man sich die Abbildungen 1.2 und 1.3. Abbildung 1.2 stellt die Approximationskurve und eine vorgegebene Punktmenge dar, die vor der Minimierung von \bar{F} skaliert wurde (nach der Minimierung wurden die resultierende Kurve und die Punkte zurückskaliert). Abbildung 1.3 zeigt die Approximation der selben Punktmenge ohne Skalierung. Beide Kurven haben die gleiche Ordnung und Segmentanzahl und interpolieren Anfangs- und Endvorgabepunkt. In beiden Abbildungen sind Fehlervektoren als „Nadeln“ eingezeichnet, die die Punkte \mathbf{P}_i mit den entsprechenden Kurvenpunkten $\bar{\mathbf{X}}(t_i)$ am Ende der Minimierung von \bar{F} (in Verbindung mit Parameterkorrektur s. u.) verbinden. In Abbildung 1.2 sind die Fehlervektoren ungefähr senkrecht zur Kurve, in Abbildung 1.3 die sehr viel längeren Vektoren im allgemeinen dagegen nicht. In beiden Fällen jedoch sind die maximalen Fehlerwerte der nicht-euklidischen Zielfunktion \bar{F} beinahe gleich.

Zum Abschluß muß noch der Gewichtskontrollterm G festgelegt werden. G soll sicherstellen, daß positive Gewichte h_j als Lösungen der Minimierung von (1.1) erzeugt werden können. Dabei soll die Linearität des Optimierungsproblems erhalten bleiben. Die Forderung nach Gewichten größer als Null gewährleistet, daß sich schöne und wichtige geometrische Eigenschaften der integralen B-Splinekurven, z. B. konvexe Hülle-Eigenschaft, auf rationale B-Splinekurven übertragen. Dazu wird G eingeführt als

$$G := \sum_{i=0}^m (x_0(t_i) - 1)^2 = \sum_{i=0}^m \left(\sum_{j=0}^n h_j N_{jk}(t_i) - 1 \right)^2$$

Die Minimierung von G allein hat die Lösungen $h_j = 1$. Der Faktor λ in (1.1) steuert den Einfluß des Kontrollterms. Bildet man nun $\partial \bar{F} / \partial h_j = 0$, d. h. stellt man einen Teil der notwendigen Bedingungsgleichungen zur Minimierung von (1.1) auf, so hat man

$$\frac{\partial F}{\partial h_j} + \lambda \frac{\partial G}{\partial h_j} = 0 \quad (1.17)$$

Dividiert man in (1.17) durch λ und bildet den Grenzwert $\lambda \rightarrow \infty$, dann erhält man $\partial G / \partial h_j = 0$ mit den Lösungen $h_j = 1$. Durch Wahl eines genügend großen λ kann daher erreicht werden, daß die Minimierung der Zielfunktion \bar{F} zu positiven Gewichten führt.

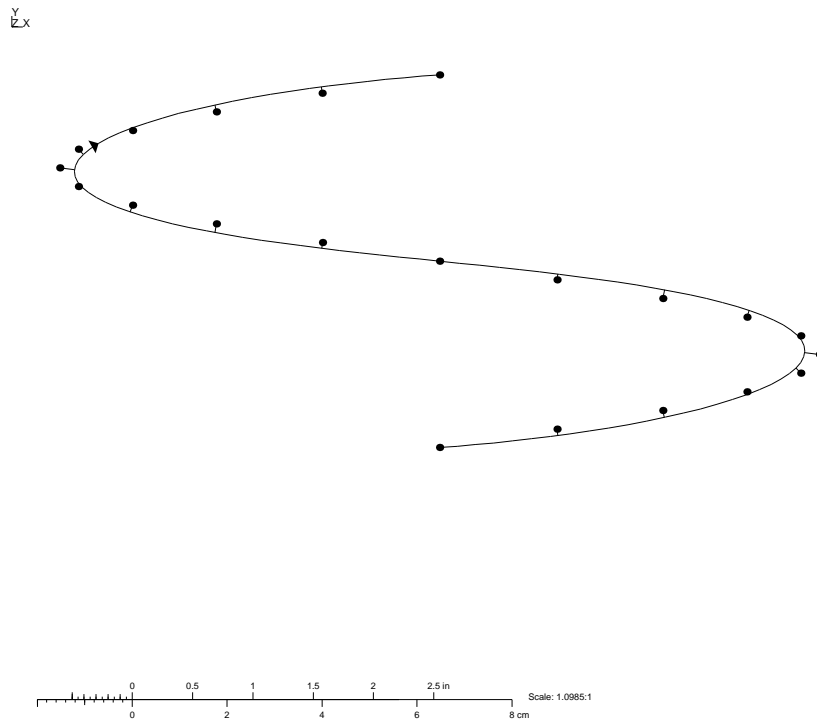


Abbildung 1.2: Approximation mit skalierten Punkten und späterer Rückskalierung

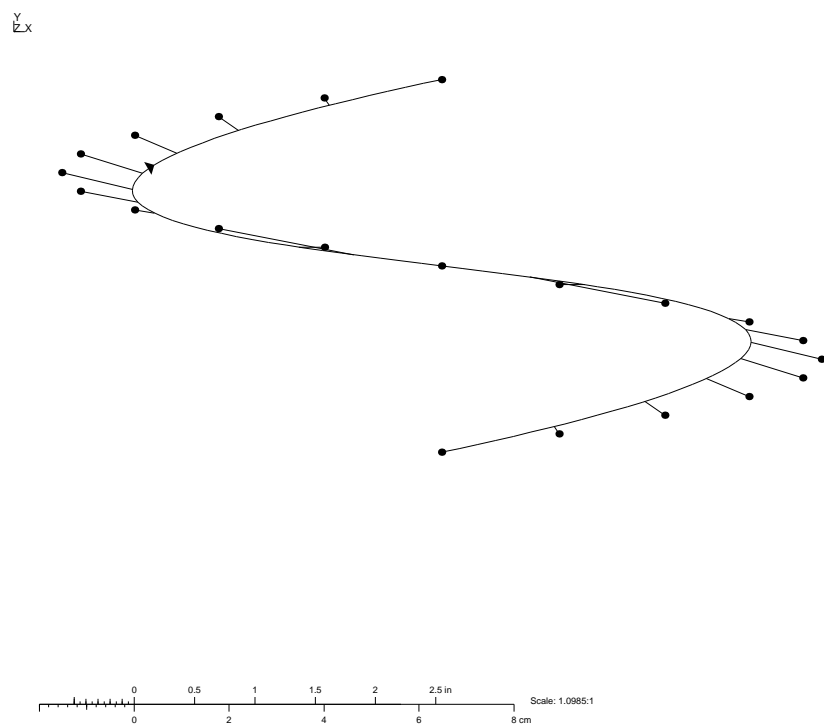


Abbildung 1.3: Approximation mit unskalierten Punkten (Punkte wie in Abbildung 1.2)

1.3 Lösen des Approximationsproblems

In den folgenden Abschnitten wird das Lösen des Approximationsproblems (Minimierung von (1.1)) näher behandelt. Zuerst wird die Existenz und Eindeutigkeit der Lösung nachgewiesen. Dann stellen wir ein Verfahren vor, mit dem in der Praxis eine Lösung bestimmt werden kann. Zum Schluß werden zusätzlich zu den de Boor-Punkten und Gewichten die Parameterwerte t_i als Optimierungsvariablen angesehen und mit Hilfe eines Iterationsverfahrens (Parameterkorrektur) wird der Wert der Zielfunktion (1.1) weiter reduziert. Aufgrund des im letzten Abschnitt Gesagten setzen wir in diesen weiteren Abschnitten voraus, daß der Schwerpunkt der Punkte \mathbf{P}_i im Koordinatenursprung liegt und $|\mathbf{P}_i| \leq 1/2$ gilt.

1.3.1 Existenz und Eindeutigkeit der Lösung

Die notwendigen Bedingungen zur Minimierung von (1.1)

$$\frac{\partial \bar{F}}{\partial \bar{d}_{jx}} = 0, \quad \frac{\partial \bar{F}}{\partial \bar{d}_{jy}} = 0, \quad \frac{\partial \bar{F}}{\partial \bar{d}_{jz}} = 0, \quad \frac{\partial \bar{F}}{\partial h_j} = 0$$

führen auf das lineare System

$$\begin{aligned} (\mathbf{A}_1 + \lambda \mathbf{A}_2) \mathbf{x} &= \lambda \mathbf{b} \\ \mathbf{A}_{11} \mathbf{x}' + (\mathbf{A}_{12} + \lambda \mathbf{A}_{22}) \mathbf{h} &= \lambda \mathbf{b} \end{aligned} \tag{1.18}$$

mit

$$\begin{aligned} \mathbf{A}_1 &= \left(\begin{array}{ccc} \sum \bar{N}_{00}^i (\mathbf{I} - c_i \mathbf{P}_i \mathbf{P}_i^\top) & \cdots & \sum \bar{N}_{0n}^i (\mathbf{I} - c_i \mathbf{P}_i \mathbf{P}_i^\top) \\ \vdots & \ddots & \vdots \\ \sum \bar{N}_{n0}^i (\mathbf{I} - c_i \mathbf{P}_i \mathbf{P}_i^\top) & \cdots & \sum \bar{N}_{nn}^i (\mathbf{I} - c_i \mathbf{P}_i \mathbf{P}_i^\top) \\ - \sum c_i \bar{N}_{00}^i \mathbf{P}_i^\top & \cdots & - \sum c_i \bar{N}_{0n}^i \mathbf{P}_i^\top \\ \vdots & \ddots & \vdots \\ - \sum c_i \bar{N}_{n0}^i \mathbf{P}_i^\top & \cdots & - \sum c_i \bar{N}_{nn}^i \mathbf{P}_i^\top \end{array} \right) \mathbf{A}_{12} \in \mathbb{R}^{4n+4, 4n+4} \\ &\quad \underbrace{\hspace{10em}}_{\mathbf{A}_{11} \in \mathbb{R}^{4n+4, 3n+3}} \\ \mathbf{A}_{12} &= \left(\begin{array}{ccc} - \sum c_i \bar{N}_{00}^i \mathbf{P}_i & \cdots & - \sum c_i \bar{N}_{0n}^i \mathbf{P}_i \\ \vdots & \ddots & \vdots \\ - \sum c_i \bar{N}_{n0}^i \mathbf{P}_i & \cdots & - \sum c_i \bar{N}_{nn}^i \mathbf{P}_i \\ \sum (1 - c_i) \bar{N}_{00}^i & \cdots & \sum (1 - c_i) \bar{N}_{0n}^i \\ \vdots & \ddots & \vdots \\ \sum (1 - c_i) \bar{N}_{n0}^i & \cdots & \sum (1 - c_i) \bar{N}_{nn}^i \end{array} \right) \in \mathbb{R}^{4n+4, n+1} \end{aligned}$$

$$\begin{aligned}
\mathbf{A}_2 &= \left(\begin{array}{ccc|c} \mathbf{0} & \cdots & \mathbf{0} & \\ \vdots & \ddots & \vdots & \\ \mathbf{0} & \cdots & \mathbf{0} & \\ 0 & \cdots & 0 & \\ \vdots & \ddots & \vdots & \\ 0 & \cdots & 0 & \end{array} \right) \mathbf{A}_{22} \in \mathbb{R}^{4n+4, 4n+4}, \quad \mathbf{A}_{22} = \left(\begin{array}{ccc|c} \mathbf{0} & \cdots & \mathbf{0} & \\ \vdots & \ddots & \vdots & \\ \mathbf{0} & \cdots & \mathbf{0} & \\ \hline \sum \tilde{N}_{00}^i & \cdots & \sum \tilde{N}_{0n}^i & \\ \vdots & \ddots & \vdots & \\ \sum \tilde{N}_{n0}^i & \cdots & \sum \tilde{N}_{nn}^i & \end{array} \right) \in \mathbb{R}^{4n+4, n+1} \\
\mathbf{x} &= (\underbrace{\bar{\mathbf{d}}_0, \dots, \bar{\mathbf{d}}_n}_{\mathbf{x}'}, \underbrace{h_0, \dots, h_n}_{\mathbf{h}})^\top \\
\mathbf{b} &= (\mathbf{0}, \dots, \mathbf{0}, \sum N_{0k}(t_i), \dots, \sum N_{nk}(t_i))^\top \\
\bar{\mathbf{d}}_j &= (\bar{d}_{jx}, \bar{d}_{jy}, \bar{d}_{jz})^\top, \quad \tilde{N}_{lj}^i = N_{lk}(t_i) N_{jk}(t_i) \\
c_i &= \frac{1}{1 + \mathbf{P}_i^2} \quad \text{und der Einheitsmatrix } \mathbf{I} \in \mathbb{R}^{3,3}
\end{aligned}$$

Die ersten $3n + 3$ Zeilen der Matrix \mathbf{A}_1 entstehen durch Ableiten der Abstandsfunktion F nach den Unbekannten $\bar{\mathbf{d}}_j$, die letzten $n + 1$ Zeilen durch Ableiten von F nach den Unbekannten h_j . Dabei sind \mathbf{A}_{11} und \mathbf{A}_{12} Teilmatrizen von \mathbf{A}_1 , deren Elemente Koeffizienten der Unbekannten $\bar{\mathbf{d}}_j$ bzw. der h_j sind. Die Zeilen der Matrix \mathbf{A}_2 ergeben sich durch Differenzieren des Gewichtskontrollterms G nach den $\bar{\mathbf{d}}_j$ und den h_j . Auch hier ist \mathbf{A}_{22} eine Teilmatrix von \mathbf{A}_2 und die Elemente von \mathbf{A}_{22} sind Koeffizienten der Unbekannten h_j . Die rechte Seite \mathbf{b} geht durch Ableiten von G nach den Variablen $\bar{\mathbf{d}}_j$ sowie h_j hervor.

Zur Lösbarkeit des Gleichungssystems (1.18) läßt sich sagen:

Satz 1.3.1 *Das lineare System (1.18) hat für festes $\lambda > 0$ und den Voraussetzungen aus Abschnitt 1.1 an die Parameterwerte t_i genau eine Lösung $\mathbf{x} \neq \mathbf{0}$, die ein lokales Minimum der Zielfunktion (1.1) ist.*

Beweis. Matrix \mathbf{A}_1 ist die zur quadratischen Form (1.14) gehörige Koeffizientenmatrix. Nach Abschnitt 1.2 ist sie positiv semidefinit. Für Matrix \mathbf{A}_2 gilt $\mathbf{x}^\top \mathbf{A}_2 \mathbf{x} > 0$, falls in \mathbf{x} wenigstens eine Gewichtskomponente $h_j \neq 0$ ist. Denn die Teilmatrix in \mathbf{A}_{22} (Matrix unterhalb des horizontalen Strichs) ist positiv definit und sonst sind alle Matrixelemente von \mathbf{A}_2 gleich Null. Sind im Vektor \mathbf{x} aber alle Gewichte $h_j = 0$, dann ist die Interpolation der Punkte \mathbf{P}_i durch die rationale B-Splinekurve nicht möglich. Das heißt aber, daß man $\mathbf{x}^\top \mathbf{A}_1 \mathbf{x} = 0$ genau dann hat, wenn $\mathbf{x} = \mathbf{0}$ ist. Zusammenfassend gilt somit $\mathbf{x}^\top (\mathbf{A}_1 + \lambda \mathbf{A}_2) \mathbf{x} > 0 \quad \forall \mathbf{x} \neq \mathbf{0}, \mathbf{x} \in \mathbb{R}^{4n+4}$. Die Matrix $\mathbf{A}_1 + \lambda \mathbf{A}_2$ ist demnach positiv definit. ■

1.3.2 Bestimmung einer Lösung

Es sollen positive Gewichte h_j berechnet werden. Daher wird von den h_j als Teil des Lösungsvektors \mathbf{x} gefordert

$$\begin{aligned}
h_j &\in [h_{\min}, h_{\max}] & (j = 0, \dots, n) \\
h_{\min} &:= 1 - c, \quad h_{\max} := 1 + c, \quad 0 < c < 1, \quad c = \text{const.}
\end{aligned} \tag{1.19}$$

Da von vornherein unklar ist, welche Größenordnung λ haben muß, um die Bedingungen (1.19) einzuhalten, muß λ im Verlauf des Lösungsverfahrens geeignet geschätzt werden. Dabei sollte für λ ein möglichst kleiner Wert gefunden werden, damit der Einfluß des Gewichtskontrollterms in (1.1) gering bleibt. Das Lösungsverfahren berücksichtigt diese Forderungen wie folgt: Das Gleichungssystem (1.18) wird umgeformt in

$$\begin{aligned} \mathbf{A}_{11}\mathbf{x}' &= -(\mathbf{A}_{12} + \lambda\mathbf{A}_{22})\mathbf{h} + \lambda\mathbf{b} \\ \mathbf{x}' &= (\bar{\mathbf{d}}_0, \dots, \bar{\mathbf{d}}_n)^\top \\ \mathbf{h} &= (h_0, \dots, h_n)^\top \end{aligned} \quad (1.20)$$

Im nächsten Schritt wird das *Householder-Verfahren* zur Lösung linearer Gleichungssysteme auf (1.20) angewendet. Das Householder-Verfahren transformiert eine Matrix $\mathbf{A} \in \mathbb{R}^{m,n}$ ($m \geq n$) mit Rang n durch Multiplikation mit einer orthogonalen Matrix \mathbf{Q}^\top in eine Matrix, bei der die ersten n Zeilen eine obere Dreiecksmatrix bilden und alle anderen Zeilen Nullzeilen sind. Somit erhält man

$$\begin{aligned} \mathbf{Q}^\top \mathbf{A}_{11}\mathbf{x}' &= \mathbf{Q}^\top (-(\mathbf{A}_{12} + \lambda\mathbf{A}_{22})\mathbf{h} + \lambda\mathbf{b}) \\ \mathbf{Q}^\top \mathbf{A}_{11} &= \begin{pmatrix} \mathbf{U} \\ \mathbf{0} \end{pmatrix}, \quad \mathbf{Q}^\top = \begin{pmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{pmatrix} \end{aligned} \quad (1.21)$$

\mathbf{Q} ist eine orthogonale Matrix und $\mathbf{U} \in \mathbb{R}^{3n+3, 3n+3}$ ist eine obere Dreiecksmatrix. Aus (1.21) folgt das lineare System

$$\mathbf{Q}_2(\mathbf{A}_{12} + \lambda\mathbf{A}_{22})\mathbf{h} = \lambda\mathbf{Q}_2\mathbf{b} \quad (1.22)$$

In (1.22) treten nur noch die Unbekannten h_j und λ auf. Falls die Matrix $\mathbf{Q}_2\mathbf{A}_{12}$ invertierbar ist, kann das Gleichungssystem umgeschrieben werden zu

$$\mathbf{h} = \lambda((\mathbf{Q}_2\mathbf{A}_{12})^{-1}\mathbf{Q}_2\mathbf{b} - (\mathbf{Q}_2\mathbf{A}_{12})^{-1}\mathbf{Q}_2\mathbf{A}_{22}\mathbf{h}) \quad (1.23)$$

Von der Matrix $\mathbf{Q}_2\mathbf{A}_{12}$ kann in vielen Fällen gezeigt werden, daß sie eine Inverse besitzt. Denn aus der Invertierbarkeit der Matrix \mathbf{A}_1 folgt die Invertierbarkeit der Matrix $\mathbf{Q}_2\mathbf{A}_{12}$. Eine hinreichende Bedingung für die Existenz der Inversen von \mathbf{A}_1 ist nach Abschnitt 1.2, daß die Interpolation aller gegebenen Punkte durch die rationale B-Splinekurve nicht möglich ist. Betragsbildung bei den Matrix- und Vektorelementen und Anwendung der Dreiecksungleichung auf (1.23) ergibt die Ungleichungen

$$|\mathbf{h}| \leq \lambda(|(\mathbf{Q}_2\mathbf{A}_{12})^{-1}\mathbf{Q}_2\mathbf{b}| + |(\mathbf{Q}_2\mathbf{A}_{12})^{-1}\mathbf{Q}_2\mathbf{A}_{22}| \cdot |\mathbf{h}|)$$

und weiter abgeschätzt

$$|\mathbf{h}_{min}| \leq \lambda(|(\mathbf{Q}_2\mathbf{A}_{12})^{-1}\mathbf{Q}_2\mathbf{b}| + |(\mathbf{Q}_2\mathbf{A}_{12})^{-1}\mathbf{Q}_2\mathbf{A}_{22}| \cdot |\mathbf{h}_{max}|) \quad (1.24)$$

Betragstriche bei den Matrizen und Vektoren sind elementweise zu verstehen. Der Vektor $\mathbf{h}_{min} := (h_{min}, \dots, h_{min})^\top$ und \mathbf{h}_{max} ist entsprechend definiert. Die einzelnen Ungleichungen

von (1.24) werden nach λ aufgelöst. Die größte dieser unteren Schranken von λ wählen wir als Schätzwert λ_s für λ . Gilt also $\lambda < \lambda_s$, dann sind die Bedingungen (1.19) nicht alle erfüllt. λ_s wird jetzt in (1.22) für λ eingesetzt und das Gleichungssystem gelöst. Die Lösungen h_j werden auf die Einhaltung der Bedingungen (1.19) geprüft. Wenn das nicht der Fall ist, wird der Vektor $\mathbf{h}'_{min} := \mathbf{h}_{min} + \delta\mathbf{h}$, $\delta\mathbf{h} = (0.1, \dots, 0.1)^T$ gebildet. $(0.1, \dots, 0.1)^T$ stellt eine empirische Wahl dar, die sich in den Testbeispielen als guter Kompromiß zwischen einer zu feinen und einer zu groben Abschätzung für λ erwiesen hat. Mit \mathbf{h}'_{min} anstelle von \mathbf{h}_{min} wird aus (1.24) ein neuer Schätzwert λ_s bestimmt. Mit dem Schätzwert wird wieder (1.22) gelöst und für die Lösungen die Bedingungen (1.19) überprüft. Sind noch nicht alle Bedingungen erfüllt, werden die einzelnen Schritte mit $\delta\mathbf{h} = (0.2, \dots, 0.2)^T$ wiederholt. Das Verfahren wird so lange durchlaufen, bis die Bedingungen (1.19) eingehalten werden. In den Testbeispielen hat sich gezeigt, daß nur wenige Iterationen notwendig sind, um λ_s geeignet zu bestimmen. Danach wird der zuletzt ermittelte Schätzwert λ_s und der dazugehörige Lösungsvektor \mathbf{h} in (1.21) eingesetzt und das Restgleichungssystem

$$\mathbf{U}\mathbf{x}' = \mathbf{Q}_1(-(\mathbf{A}_{12} + \lambda\mathbf{A}_{22})\mathbf{h} + \lambda\mathbf{b}) \quad (1.25)$$

gelöst.

Bei der praktischen Anwendung des eben beschriebenen Verfahrens sollte darauf geachtet werden, daß die Anzahl der Approximationspunkte und die Anzahl der Kontrollpunkte die Beziehung $n < 3/4 m$ bei Raumkurven und $n < 2/3 m$ bei ebenen Kurven einhält. $3/4 m$ im räumlichen Fall und $2/3 m$ im ebenen Fall sind empirisch gefundene obere Schranken, die bei Überschreitung mit n dazu führen können, daß Lücken im linearen System (1.18) auftreten. Weiterhin sollten die Approximationspunkte gleichmäßig derart verteilt sein, daß die Anzahl der den Punkten \mathbf{P}_i zugeordneten Parameterwerte t_i , die die Ungleichungen $\tau_j < t_i < \tau_{j+k}$ ($j = 0, \dots, n$) erfüllen, in etwa gleich ist. Ansonsten kann die erzeugte rationale Approximationskurve zu Oszillationen neigen, was unerwünscht ist.

Das beschriebene Lösungsverfahren hat für die praktische Anwendung eine erwünschte Eigenschaft:

Satz 1.3.2 *Werden die Punkte \mathbf{P}_i mit einer orthogonalen Drehmatrix \mathbf{D} transformiert, dann liefert das Verfahren (1.20)-(1.25) die Lösungen*

$$\tilde{\mathbf{x}} = (\mathbf{D}\tilde{\mathbf{d}}_0, \dots, \mathbf{D}\tilde{\mathbf{d}}_n, h_0, \dots, h_n)^T$$

Dabei sind $\tilde{\mathbf{d}}_j$, h_j die mit demselben Verfahren berechneten Lösungen bei der Approximation der untransformierten Punkte \mathbf{P}_i mit den gleichen Parameterwerten t_i .

Beweis. Das Gleichungssystem mit den transformierten Punkten kann mit den Bezeichnungen aus (1.18) geschrieben werden als

$$\tilde{\mathbf{D}}(\mathbf{A}_1 + \lambda\mathbf{A}_2)\tilde{\mathbf{D}}^T \tilde{\mathbf{x}} = \lambda\tilde{\mathbf{D}}\mathbf{b} \quad (1.26)$$

Die orthogonale Matrix $\tilde{\mathbf{D}}$ ist gegeben durch

$$\tilde{\mathbf{D}} = \begin{pmatrix} \mathbf{D} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix}$$

mit der Einheitsmatrix $\mathbf{I} \in \mathbb{R}^{n+1, n+1}$. Durch Auflösen des Systems (1.26) nach $\tilde{\mathbf{x}}$ erhält man

$$\tilde{\mathbf{x}} = \tilde{\mathbf{D}}\mathbf{x} \quad ,$$

wenn \mathbf{x} die Lösung von (1.18) ist. Der Satz ist somit bewiesen, falls das Verfahren im Transformationsfall denselben Schätzwert λ_s bestimmt. Dem Gleichungssystem (1.22) entspricht im Transformationsfall das System

$$\tilde{\mathbf{Q}}_2 \tilde{\mathbf{D}} (\mathbf{A}_{12} + \lambda \mathbf{A}_{22}) \mathbf{h} = \lambda \tilde{\mathbf{Q}}_2 \tilde{\mathbf{D}} \mathbf{b}$$

Zu zeigen ist demnach noch, daß gilt

$$(\tilde{\mathbf{Q}}_2 \tilde{\mathbf{D}} \mathbf{A}_{12})^{-1} \tilde{\mathbf{Q}}_2 \tilde{\mathbf{D}} \mathbf{w} = (\mathbf{Q}_2 \mathbf{A}_{12})^{-1} \mathbf{Q}_2 \mathbf{w} \quad \forall \mathbf{w} \in \mathbb{R}^{4n+4}$$

Dazu wird $\hat{\mathbf{Q}}_2 := \tilde{\mathbf{Q}}_2 \tilde{\mathbf{D}}$ gesetzt und die Bezeichnungen

$$\begin{aligned} \ker \mathbf{A} &= \{ \mathbf{w} \in \mathbb{R}^m \mid \mathbf{A} \mathbf{w} = \mathbf{0} \} \\ \text{im } \mathbf{A} &= \{ \mathbf{y} \in \mathbb{R}^l \mid \exists \mathbf{w} \in \mathbb{R}^m \text{ mit } \mathbf{y} = \mathbf{A} \mathbf{w} \} \end{aligned}$$

für eine Matrix $\mathbf{A} \in \mathbb{R}^{l, m}$ eingeführt. Mit Hilfe der linearen Algebra lassen sich folgende Aussagen leicht zeigen: Wenn die Matrix $\mathbf{Q}_2 \mathbf{A}_{12}$ invertierbar ist, dann muß die zur Matrix $\mathbf{A}_{12} \in \mathbb{R}^{4n+4, n+1}$ gehörige lineare Abbildung injektiv und die zur Matrix $\mathbf{Q}_2 \in \mathbb{R}^{n+1, 4n+4}$ gehörige lineare Abbildung surjektiv sein. Außerdem muß

$$\begin{aligned} \text{im } \mathbf{A}_{12} \cap \ker \mathbf{Q}_2 &= \{ \mathbf{0} \} \quad \text{und} \\ \text{im } \mathbf{A}_{12} + \ker \mathbf{Q}_2 &= \mathbb{R}^{4n+4} \end{aligned} \tag{1.27}$$

sein. Entsprechend gilt für die Matrix $\hat{\mathbf{Q}}_2 \mathbf{A}_{12}$

$$\begin{aligned} \text{im } \mathbf{A}_{12} \cap \ker \hat{\mathbf{Q}}_2 &= \{ \mathbf{0} \} \quad \text{und} \\ \text{im } \mathbf{A}_{12} + \ker \hat{\mathbf{Q}}_2 &= \mathbb{R}^{4n+4} \end{aligned} \tag{1.28}$$

Aus (1.27), (1.28) folgt sofort die Identität $\ker \mathbf{Q}_2 = \ker \hat{\mathbf{Q}}_2$. Kommen wir wieder auf die Behauptung zurück und formulieren sie um:

$$\begin{aligned} \mathbf{Q}_2 \mathbf{A}_{12} \mathbf{y} &= \mathbf{Q}_2 \mathbf{w} \\ \hat{\mathbf{Q}}_2 \mathbf{A}_{12} \tilde{\mathbf{y}} &= \hat{\mathbf{Q}}_2 \mathbf{w} \end{aligned} \tag{1.29}$$

Zu beweisen ist, daß $\mathbf{y} = \tilde{\mathbf{y}}$ gilt.

(1.29) \iff

$$\begin{aligned}\mathbf{Q}_2(\mathbf{A}_{12}\mathbf{y} - \mathbf{w}) &= \mathbf{0} \\ \hat{\mathbf{Q}}_2(\mathbf{A}_{12}\tilde{\mathbf{y}} - \mathbf{w}) &= \mathbf{0}\end{aligned}$$

\iff

$$\begin{aligned}\mathbf{A}_{12}\mathbf{y} &= \mathbf{w} + \mathbf{z} \quad , \quad \mathbf{z} \in \ker \mathbf{Q}_2 \\ \mathbf{A}_{12}\tilde{\mathbf{y}} &= \mathbf{w} + \tilde{\mathbf{z}} \quad , \quad \tilde{\mathbf{z}} \in \ker \hat{\mathbf{Q}}_2\end{aligned}$$

\implies

$$\mathbf{A}_{12}(\mathbf{y} - \tilde{\mathbf{y}}) = \mathbf{z} - \tilde{\mathbf{z}}$$

Wegen $\ker \mathbf{Q}_2 = \ker \hat{\mathbf{Q}}_2$, (1.27) und $\ker \mathbf{A}_{12} = \{\mathbf{0}\}$ folgt $\mathbf{y} = \tilde{\mathbf{y}}$. \blacksquare

Der letzte Satz besagt, daß das Lösungsverfahren invariant gegenüber Drehungen des E^3 ist, d. h. für Punkte in gedrehter Lage wird die bis auf die Drehung gleiche rationale B-Splinekurve berechnet.

1.3.3 Parameterkorrektur

Die Parameterwerte t_i zu den gegebenen Punkten \mathbf{P}_i legen neben den Optimierungsvariablen h_j , \bar{d}_{jx} , \bar{d}_{jy} und \bar{d}_{jz} den Wert der Zielfunktion (1.1) fest. Faßt man die Parameterwerte t_i zusätzlich als Optimierungsvariablen auf, so kann der Zielfunktionswert noch verringert werden. Zur Vermeidung eines nichtlinearen Optimierungsproblems ist die Idee der Parameterkorrektur, vgl. [11], entwickelt worden. Hierbei wird die Zielfunktion als Funktion der Parameterwerte

$$\bar{F} = \bar{F}(t_0, \dots, t_m)$$

betrachtet. Gesucht werden korrigierte Parameterwerte t'_i als Lösungen des Optimierungsproblems

$$\bar{F}(t'_0, \dots, t'_m) \longrightarrow \min \quad (1.30)$$

Bei der Lösung von (1.30) sollen die mit dem Verfahren (1.20)-(1.25) ermittelte rationale B-Splinekurve und λ unverändert bleiben. Die Zielfunktion (1.1) besteht aus $m+1$ Summanden $D(\bar{\mathbf{X}}(t_i), \bar{\mathbf{P}}_i)$. Jeder dieser Summanden hängt von genau einem Parameterwert t_i ab. Das Optimierungsproblem (1.30) kann daher auf die $m+1$ Optimierungsprobleme

$$D(\bar{\mathbf{X}}(t'_i), \bar{\mathbf{P}}_i) \longrightarrow \min \quad (i = 0, \dots, m)$$

von einer Variablen zurückgeführt werden. Korrigierte Parameterwerte können näherungsweise zum Beispiel mit dem *eindimensionalen Newton-Verfahren* berechnet werden.

Mit den korrigierten Parameterwerten t'_i kann jetzt wieder eine neue Approximationskurve durch Lösen von (1.18) bestimmt werden. Die abwechselnde Durchführung von Approximation und Parameterkorrektur liefert eine Folge von Zielfunktionswerten. Die Konvergenz der Folge gegen ein Minimum der Zielfunktion hängt von den Startparameterwerten t_i ab, da es sich um ein lokales Optimierungsverfahren handelt. Im Unterschied zu den vorherigen Abschnitten sind auch noch die Parameterwerte Optimierungsvariablen der Zielfunktion. Geeignete Startwerte für die Parameterwerte t_i erhält man zum Beispiel durch chordale oder zentripetale Parametrisierung, s. [11], der Approximationspunkte \mathbf{P}_i .

1.4 Beispiele

In den Abbildungen der nächsten Seiten sind die Vorgabepunkte und die dazugehörigen Approximationskurven gezeichnet. Zu einem Punktdatensatz werden zum Vergleich jeweils zwei Bilder gezeigt. Im oberen Bild ist die Approximationskurve dargestellt, die mit dem Verfahren aus Abschnitt 1.3 erzeugt wurde. Im unteren Bild sieht man eine Approximationskurve, die mit dem Verfahren von [25] entstand. In allen Beispielen wurden offene kubische B-Splinekurven ($k = 4$) verwendet. Der Knotenvektor \mathbf{T} wurde immer äquidistant mit einfachen inneren Knoten gewählt. In der Beschreibung zu den Bildern sind die maximalen Lotfehler der Kurven zu den Punkten in mm angegeben. Zum Einschätzen der Größenverhältnisse zwischen den Kurven und den Fehlerwerten sind Maßstäbe in die Bilder eingezeichnet. Zusätzlich sind in Richtung der Normalen Krümmungswerte als „Nadeln“ an die Kurve angetragen. Die Nadel ist um so länger, je größer die Kurvenkrümmung ist. Desweiteren ist der Durchlaufsinne der Kurven durch Dreiecke markiert (der Eckpunkt des Dreiecks auf der Kurve zeigt die Richtung an).

Im ersten Beispiel ist die Approximation von 25 symmetrisch auf einem Quadrat gelegenen Vorgabepunkten dargestellt. Der Fehler der oberen Kurve beträgt 1.92 mm, der Fehler der unteren Kurve 1.84 mm und beide Kurven haben 4 Segmente. Das Berechnungsverfahren für die obere Kurve benötigte bei 60 Iterationen eine Zeit¹ von 1.5 s, das für die untere Kurve bei 20 Iterationen 5.3 s. Man sieht, daß beide Kurven mit Anfangs- und Endpunkt in der linken oberen Ecke die Symmetrie der Punkte erhalten, die obere Kurve aber die linke untere Ecke und die rechte obere Ecke des Quadrats „runder“ approximiert.

Als zweites Beispiel sind wiederum auf einem Quadrat vorgegebene Punkte approximiert. Diesmal jedoch sind es wesentlich mehr (105) und dichter liegende Punkte. Die obere Kurve mit 8 Segmenten hat einen Fehler von 0.32 mm bei einer Rechenzeit von 16.8 s (90 Iterationen), die untere Kurve mit gleicher Segmentanzahl einen Fehler von 0.67 mm bei einer Rechenzeit von 3 min 22.8 s (20 Iterationen). Ähnlich wie im ersten Beispiel bewahren beide Kurven mit Anfangs- und Endpunkt in der linken oberen Ecke die Symmetrie der Vorgabepunkte. Im Unterschied zu Beispiel 1 werden jetzt aber die Ecken besser approximiert, was besonders bei der oberen Kurve deutlich wird. Weiterhin ist die Rechenzeit für die obere Kurve bei halben Fehler um etwa den Faktor 12 kürzer.

Abbildung 1.6 zeigt Approximationskurven von 21 herzförmig gelegenen Vorgabepunkten. Die Kurven haben jeweils 8 Segmente. Die Rechenzeit für die obere Kurve betrug bei einem Fehler von 0.59 mm und 3 Iterationen 0.3 s, die Rechenzeit für die untere Kurve bei einem Fehler von 0.67 mm und 20 Iterationen 6.2 s. Hier fällt besonders die kürzere Rechenzeit für die obere Kurve auf (Faktor 20!).

Im nächsten Beispiel sind 41 Punkte auf dem Buchstaben M vorgegeben, die von rationalen B-Splinekurven mit 6 Segmenten approximiert werden. Der Fehler der oberen Kurve liegt bei 0.55 mm, der Fehler der unteren bei 0.70 mm. Die Berechnung der oberen Kurve brauchte für 40 Iterationen 2.6 s, die Berechnung der unteren Kurve brauchte für 20 Iterationen 16.6 s. Es zeigt sich, daß beide Kurven insbesondere die obere geradlinige Bereiche gut approximieren (kurze Nadeln) und die obere zudem eine günstigere Rechenzeit aufweist. Weiterhin bewahrt die obere Kurve die Symmetrie der Punkte. Auch die Umkehrung der

¹CPU-Zeit auf einer HP 9000-7xx Workstation

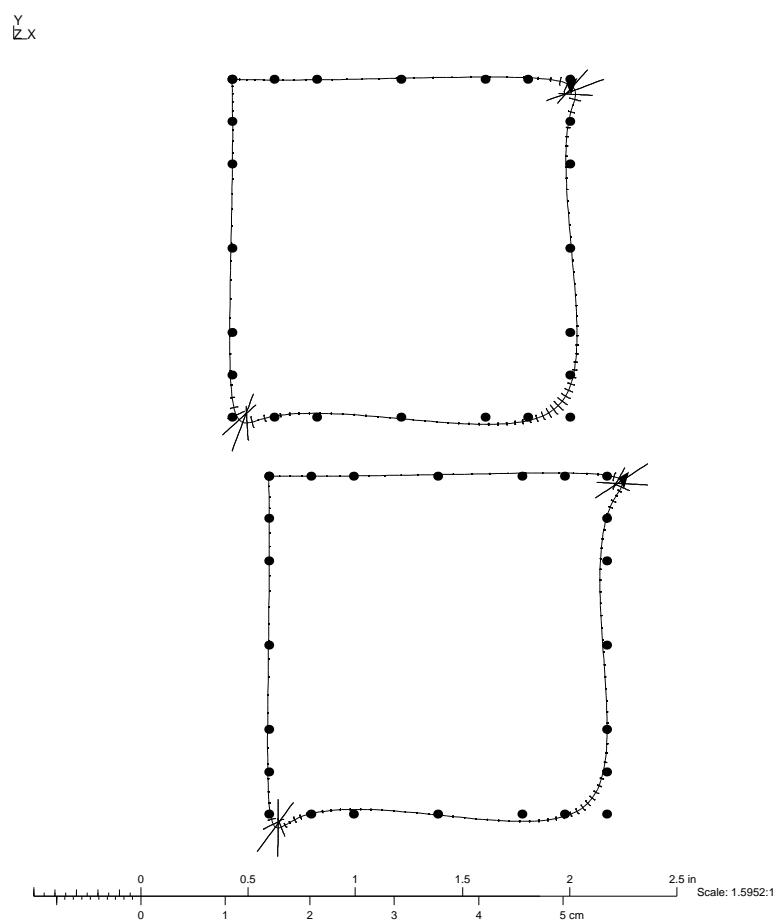


Abbildung 1.4: Approximation von quadratförmig gelegenen Punkten. Max. Fehler für die obere Kurve mit neuem Verfahren 1.92 mm, für die untere Kurve mit Verfahren [25] 1.84 mm

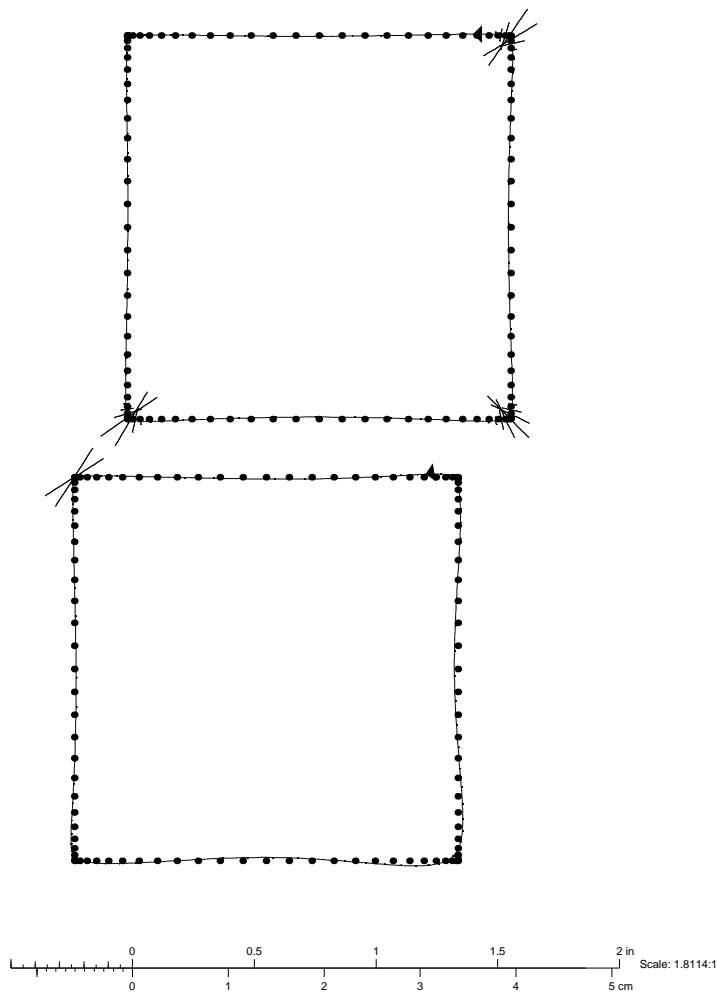


Abbildung 1.5: Approximation von quadratförmig dicht gelegenen Punkten. Max. Fehler für die obere Kurve mit neuem Verfahren 0.32 mm, für die untere Kurve mit Verfahren [25] 0.67 mm

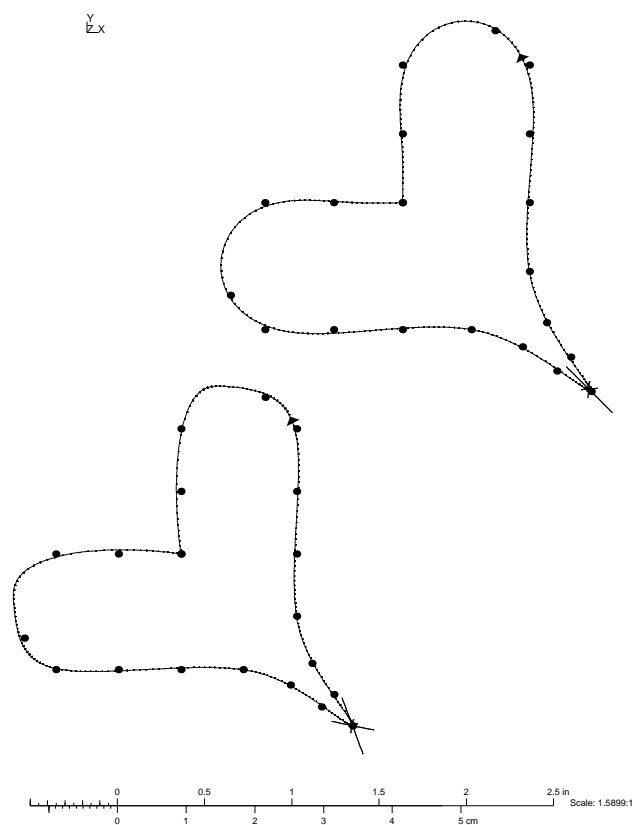


Abbildung 1.6: Approximation von herzförmig gelegenen Punkten.
 Max. Fehler für die obere Kurve mit neuem Verfahren 0.59 mm, für
 die untere Kurve mit Verfahren [25] 0.67 mm

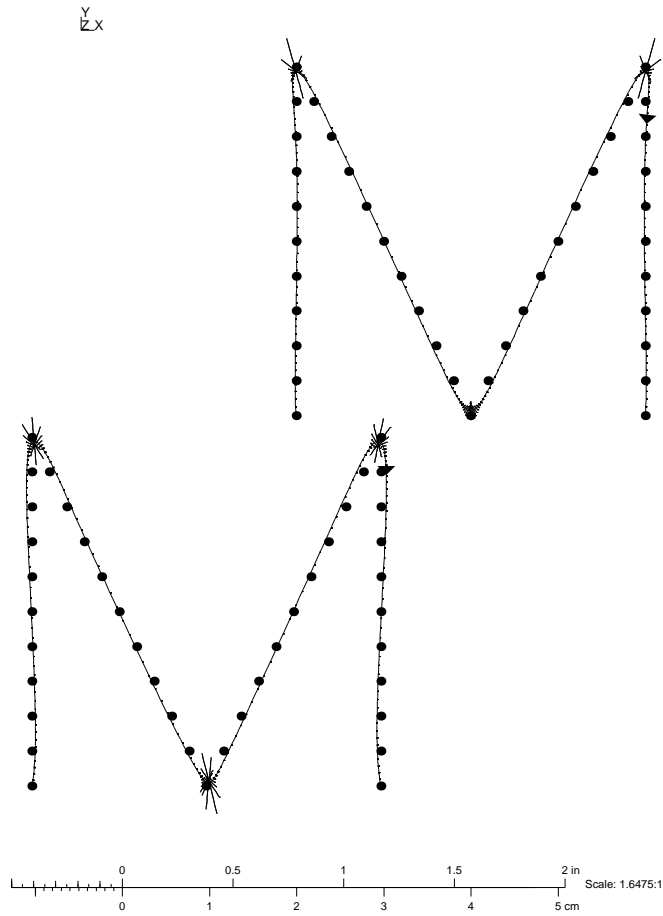


Abbildung 1.7: Approximation von M-förmig gelegenen Punkten.
 Max. Fehler für die obere Kurve mit neuem Verfahren 0.55 mm,
 für die untere Kurve mit Verfahren [25] 0.70 mm

Reihenfolge der Startparameterwerte t_i der Punkte führt bis auf Rundungsfehler auf die selbe Kurve (s. Abbildung 1.8).

Abschließend sieht man in Abbildung 1.9 sinusförmig gelegene Vorgabepunkte und zwei-segmentige Approximationskurven. Die obere Kurve hat einen Fehler von 0.09 mm und die Berechnung benötigte 0.6 s für 50 Iterationen. Die untere Kurve hat einen Fehler von 0.19

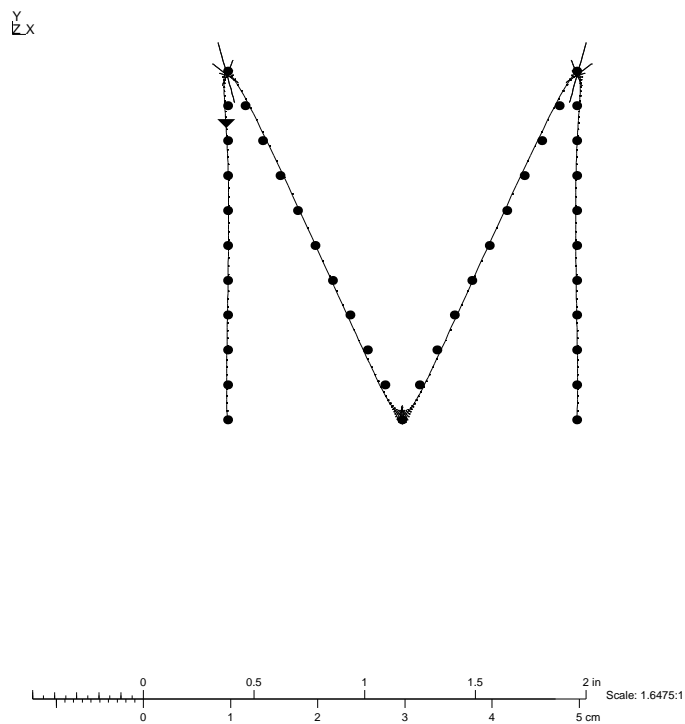


Abbildung 1.8: Approximation von M-förmig gelegenen Punkten in umgekehrter Reihenfolge (Kurve mit neuem Verfahren)

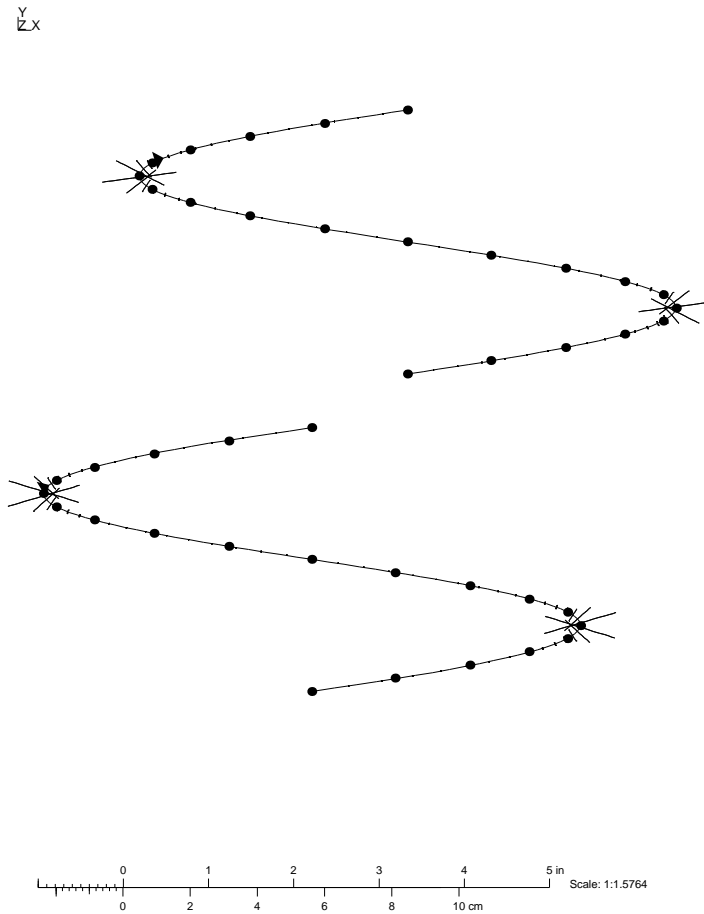


Abbildung 1.9: Approximation von sinusförmig gelegenen Punkten.
 Max. Fehler für die obere Kurve mit neuem Verfahren 0.09 mm, für
 die untere Kurve mit Verfahren [25] 0.19 mm

mm und eine Berechnungszeit von 2.8 s für 20 Iterationen. Beide Kurven unterscheiden sich in ihrem Aussehen trotz des deutlich kleineren Fehlers der oberen Kurve (Faktor 2) nur geringfügig. Wiederum wird jedoch die obere Kurve schneller berechnet.

Insgesamt besitzt das neue Approximationsverfahren folgende Merkmale:

- Das neue Approximationsverfahren erreicht bei gleicher Segmentanzahl maximale Lot-

fehler in der Größenordnung des Verfahrens von [25], teilweise sogar bessere Fehlerwerte.

- Die Rechenzeit ist dabei im Vergleich zu [25] günstiger. Besonders deutlich wird dies an den Laufzeiten zu den Abbildungen 1.5, 1.6 und 1.7.
- Der Speicherbedarf im Rechner ist beim neuen Verfahren bei gleicher Segmentanzahl meist geringer (für $m > 15$) als bei [25]. Das Verhältnis zur Speicherung der Koeffizienten der Gleichungssysteme vom neuen Verfahren zu [25] ist nämlich $16/(m+1)$. Das Verhältnis wird folglich um so kleiner je mehr Punkte approximiert werden.
- Das neue Approximationsverfahren bewahrt die Symmetrie der Punktdaten (siehe insbesondere Abbildungen 1.5, 1.7 und 1.8).
- Die Ordnung k der Kurve ist bei dem neuen Verfahren prinzipiell nicht eingeschränkt, während das Verfahren von [25] mit vertretbarem Zeitaufwand nur bis zu Ordnungen $k < 5$ anwendbar ist. Ein Beispiel ist für $k = 5$ in Abbildung 1.10 zu sehen. Hier sind die herzförmig gelegenen Punkte aus Abbildung 1.6 noch einmal approximiert worden. Die Kurve hat 7 Segmente und damit die gleiche Anzahl von de Boor-Punkten wie die Kurven in Abbildung 1.6. Der Fehler beträgt 0.62 mm (Rechenzeit 0.4 s für 5 Iterationen).

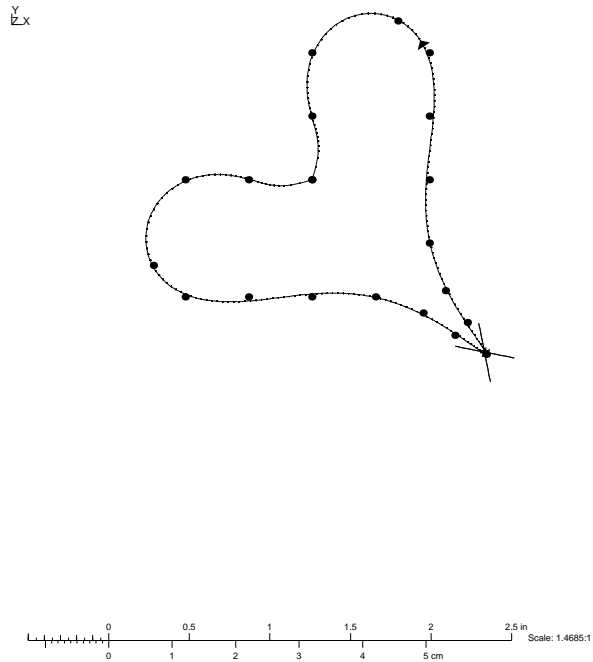


Abbildung 1.10: Approximation von herzförmig gelegenen Punkten durch eine Kurve der Ordnung $k = 5$ mit max. Fehler 0.62 mm

Kapitel 2

Approximation mit rationalen B-Splineflächen

Die Ergebnisse des letzten Kapitels werden jetzt auf rationale B-Splineflächen übertragen, um ein Approximationsverfahren für rationale Flächen zu erhalten.

2.1 Das Approximationsproblem

Gegeben sind $m + 1$ Punkte im reellen euklidischen Raum E^3 mit den kartesischen Koordinaten $\mathbf{P}_i = (p_{ix}, p_{iy}, p_{iz})^\top \in \mathbb{R}^3$ und den Parameterwerten (u_i, v_i) . Aus den in Abschnitt 1.2 genannten Gründen wird vorausgesetzt, daß der Schwerpunkt der Punktkoordinaten \mathbf{P}_i im Koordinatenursprung liegt und die \mathbf{P}_i so skaliert werden, daß $|\mathbf{P}_i| \leq 1/2$ gilt. Gesucht ist die rationale B-Splinefläche mit den Ordnungen k, p gegeben durch

$$\mathbf{X}(u, v) := \frac{\sum_{j=0}^n \sum_{l=0}^r h_{jl} \mathbf{d}_{jl} N_{jk}(u) N_{lp}(v)}{\sum_{j=0}^n \sum_{l=0}^r h_{jl} N_{jk}(u) N_{lp}(v)} \quad , \quad \mathbf{d}_{jl} := \begin{pmatrix} d_{jlx} \\ d_{jly} \\ d_{jlz} \end{pmatrix}$$

Die \mathbf{d}_{jl} sind die de Boor-Punkte und die h_{jl} die Gewichte. Die B-Splinefunktionen N_{jk} der Ordnung k und N_{lp} der Ordnung p sind dabei über den Knotenvektoren

$$\begin{aligned} \mathbf{U} &:= (\mu_0 = \dots = \mu_{k-1}, \mu_k, \dots, \mu_n, \mu_{n+1} = \dots = \mu_{n+k}) \quad , \quad n \geq k - 1 \\ \mathbf{V} &:= (\nu_0 = \dots = \nu_{p-1}, \nu_p, \dots, \nu_r, \nu_{r+1} = \dots = \nu_{r+p}) \quad , \quad r \geq p - 1 \end{aligned}$$

definiert und es gilt $(n + 1)(r + 1) < m + 1$ sowie $u_0 = \mu_0$, $u_m = \mu_{n+1}$, $v_0 = \nu_0$, $v_m = \nu_{r+1}$. Außerdem sollen die Parameterwerte (u_i, v_i) folgende Bedingungen erfüllen: es müssen Teilmengen $U_j \subset \{(u_0, v_0), \dots, (u_m, v_m)\}$ existieren, so daß gilt

$$\mu_j < u_i < \mu_{j+k} \quad \forall (u_i, v_i) \in U_j \quad (j = 0, \dots, n)$$

und $u_i < u_s \quad \forall (u_i, v_i) \in U_j, \forall (u_s, v_s) \in U_{j+1}$ und die v_i mit $(u_i, v_i) \in U_j$ genügen den *Schönberg-Whitney* Bedingungen (s. Abschnitt 1.1) bezüglich des Knotenvektors \mathbf{V} . Der

Kurvenapproximation entsprechend wählen wir für die Fläche und die Punkte die homogenen Koordinaten

$$\bar{\mathbf{X}}(u, v) := \begin{pmatrix} \sum_{j=0}^n \sum_{l=0}^r h_{jl} N_{jk}(u) N_{lp}(v) \\ \sum_{j=0}^n \sum_{l=0}^r \bar{d}_{jlx} N_{jk}(u) N_{lp}(v) \\ \sum_{j=0}^n \sum_{l=0}^r \bar{d}_{jly} N_{jk}(u) N_{lp}(v) \\ \sum_{j=0}^n \sum_{l=0}^r \bar{d}_{jlz} N_{jk}(u) N_{lp}(v) \end{pmatrix}, \quad \bar{\mathbf{P}}_i := c_i \begin{pmatrix} 1 \\ p_{ix} \\ p_{iy} \\ p_{iz} \end{pmatrix} \quad (2.1)$$

mit $\bar{d}_{jlx} := h_{jl} d_{jlx}$, $\bar{d}_{jly} := h_{jl} d_{jly}$, $\bar{d}_{jlz} := h_{jl} d_{jlz}$ und

$$c_i := \frac{1}{\sqrt{1 + p_{ix}^2 + p_{iy}^2 + p_{iz}^2}}$$

Ersetzt man in (1.1) die homogenen Kurvenkoordinaten durch (2.1), dann erhält man die Zielfunktion

$$\bar{F} = \sum_{i=0}^m \bar{\mathbf{X}}^2(u_i, v_i) - (\bar{\mathbf{X}}(u_i, v_i) \cdot \bar{\mathbf{P}}_i)^2 + \lambda \sum_{i=0}^m \left(\sum_{j=0}^n \sum_{l=0}^r h_{jl} N_{jk}(u_i) N_{lp}(v_i) - 1 \right)^2 \quad (2.2)$$

für die Flächenapproximation. Unbekannte der Optimierung sind die h_{jl} , \bar{d}_{jlx} , \bar{d}_{jly} und \bar{d}_{jlz} . Die Bedingungen zur Minimierung von (2.2) ergeben ein lineares Gleichungssystem analog zu (1.18). Für das Lösen des Gleichungssystems gelten alle Aussagen und Sätze aus Abschnitt 1.3.

2.2 Lösungsverfahren

Das Lösungsverfahren setzt sich wie bei der Kurvenapproximation aus zwei Optimierungsteilen zusammen, die iterativ durchlaufen werden. Im ersten Teil wird das bei der Minimierung von (2.2) entstehende System mit dem Verfahren (1.20)-(1.25) bei festen Parameterwerten (u_i, v_i) gelöst. Im zweiten Teil (Parameterkorrektur) wird

$$\bar{F}((u'_0, v'_0), \dots, (u'_m, v'_m)) \longrightarrow \min$$

bei unverändert bleibender rationaler B-Splinefläche und λ minimiert. Hierbei müssen jetzt $m+1$ zweidimensionale Optimierungsprobleme mit den Variablen (u'_i, v'_i) gelöst werden. Zur näherungsweisen Lösung bietet sich beispielsweise das *zweidimensionale Newton-Verfahren* an. Mit den korrigierten Parameterwerten kann nun wieder der erste Optimierungsteil durchgeführt werden. Am Ende des Lösungsverfahrens werden die Punkte und die resultierende Fläche zurückskaliert.

2.3 Beispiele

2.3.1 Rationale Approximationsverfahren im Vergleich

Das Verfahren des vorherigen Abschnitts wird verglichen mit der rationalen Flächenapproximation von [25]. Zu denselben Vorgabepunkten werden zwei Flächen dargestellt. Die

obere Approximationsfläche wurde mit dem eben beschriebenen Verfahren, die untere mit dem Verfahren von [25] erzeugt. In allen Beispielen sind bikubische Flächen ($k = p = 4$) dargestellt und es wurden jeweils äquidistante Knotenvektoren \mathbf{U} , \mathbf{V} mit einfachen inneren Knoten verwendet. Die im nachfolgenden Text angegebenen Fehler sind die maximalen Lotfehler zwischen den Flächen und den Vorgabepunkten in mm. Um die Größenverhältnisse zwischen den Flächen und den Fehlerwerten einzuordnen, sind in jeder Abbildung Maßstäbe eingetragen. Außerdem sind an den Flächen die Absolutkrümmungen in Richtung der Normalen als „Nadeln“ angetragen. Die Absolutkrümmung A ist dabei definiert als

$$A := \frac{1}{2} (|\kappa_1| + |\kappa_2|)$$

κ_1 und κ_2 sind die Hauptkrümmungen in einem Flächenpunkt. Die Nadel ist um so länger, je größer die Absolutkrümmung ist.

Im ersten Beispiel ist die Approximation von 150 auf einem Zylinder gelegenen Vorgabepunkten dargestellt. Der Fehler der oberen Fläche beträgt 0.08 mm, der Fehler der unteren Fläche 0.25 mm und beide Flächen haben 1×2 Segmente. Das Berechnungsverfahren für die obere Fläche benötigte bei 40 Iterationen eine Zeit¹ von 9.3 s, das für die untere Fläche bei 20 Iterationen 1 min 30.3 s. Man erkennt, daß die obere Fläche mit einem um den Faktor 3 kleineren Fehler den in der einen Richtung geradlinigen Verlauf der Punkte besser annähert und die dazu gehörige Rechenzeit dabei um den Faktor 9 kürzer ist.

Abbildung 2.2 zeigt Approximationsflächen von 271 auf einer Schraubfläche gelegenen Vorgabepunkten. Die Flächen haben jeweils 3×2 Segmente. Die Rechenzeit für die obere Fläche betrug bei einem Fehler von 0.38 mm und 55 Iterationen 25.3 s, die Rechenzeit für die untere Fläche bei einem Fehler von 0.36 mm und 20 Iterationen 15 min 44.8 s. Hier ist ebenfalls ein deutlicher Zeitvorteil des neuen Approximationsverfahrens gegenüber dem von [25] vorhanden (etwa Faktor 35!). Dagegen unterscheiden sich die Flächen in ihrer Form nur gering.

Im nächsten Beispiel sind 238 Punkte auf einer Translationsfläche vorgegeben, die von rationalen B-Splineflächen mit 2×2 Segmenten approximiert werden. Der Fehler der oberen Fläche liegt bei 0.12 mm und die Berechnung mit 30 Iterationen dauerte 23.9 s. Der Fehler der unteren liegt bei 0.27 mm und einer Berechnungszeit mit 20 Iterationen von 8 min 9.6 s. Beide Flächen approximieren die Punkte recht gut, wobei die obere Fläche innere Punkte teilweise besser annähert. Die Rechenzeit für die obere Fläche ist zudem noch wesentlich günstiger (Faktor 20!).

Abbildung 2.4 enthält Approximationen von 205 Punkten auf einer Würfecke. Die obere Fläche mit 4×3 Segmenten hat einen Fehler von 0.40 mm, die untere mit 3×3 Segmenten einen Fehler von 0.36 mm. Die Rechenzeit für die obere Fläche (120 Iterationen) lag bei 1 min 56.9 s, die Rechenzeit für die untere (20 Iterationen) lag bei 9 min 52.6 s. Beide Flächen approximieren die ebenen Bereiche der Punktmenge gut (kurze Nadeln). Auch die Würfelkanten werden von beiden Flächen angemessen wiedergegeben, wenn man berücksichtigt, daß die Flächen über Knotenvektoren mit einfachen inneren Knoten definiert sind.

Abschließend kann folgendes Fazit gezogen werden:

¹CPU-Zeit auf einer HP 9000-7xx Workstation

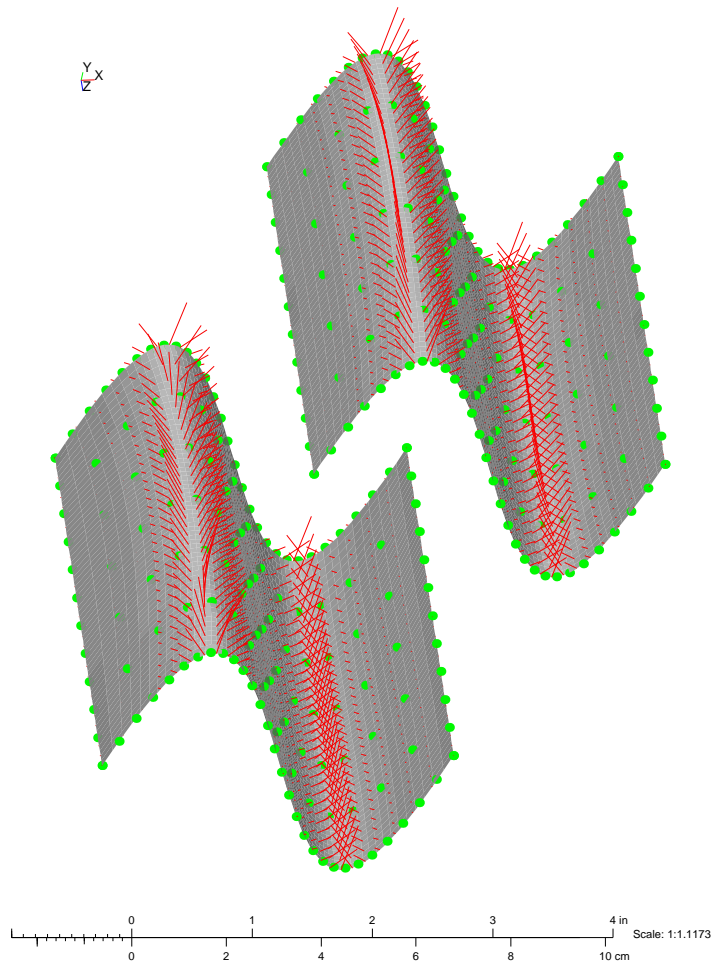


Abbildung 2.1: Approximation von Punkten auf einem Zylinder.
 Max. Fehler für die obere Fläche mit neuem Verfahren 0.08 mm,
 für die untere Fläche mit Verfahren [25] 0.25 mm

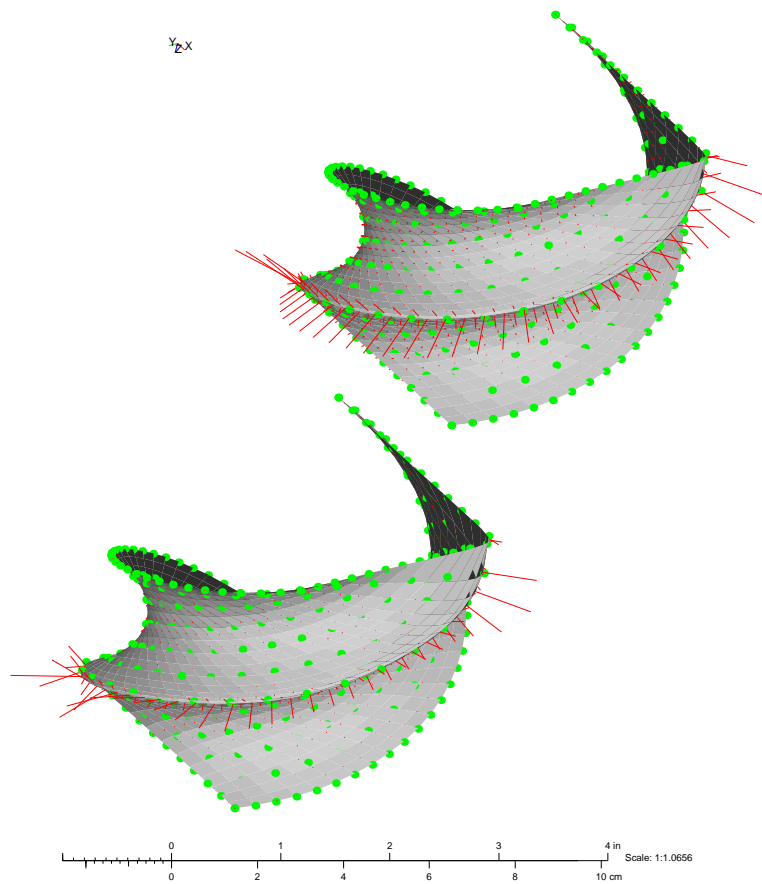


Abbildung 2.2: Approximation von Punkten auf einer Schraubfläche. Max. Fehler für die obere Fläche mit neuem Verfahren 0.38 mm, für die untere Fläche mit Verfahren [25] 0.36 mm

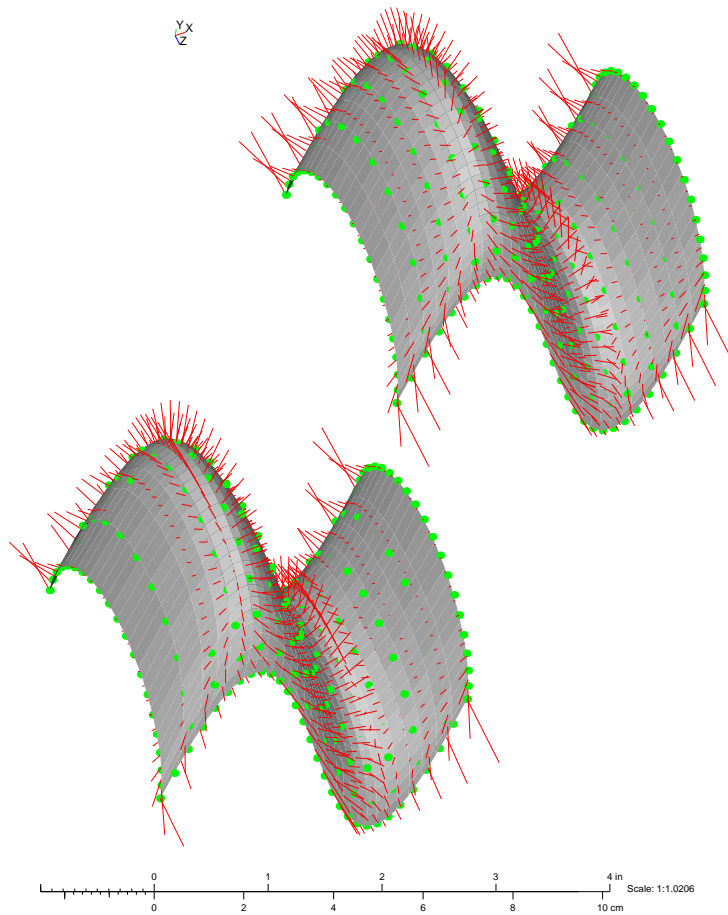


Abbildung 2.3: Approximation von Punkten auf einer Translationsfläche. Max. Fehler für die obere Fläche mit neuem Verfahren 0.12 mm, für die untere Fläche mit Verfahren [25] 0.27 mm

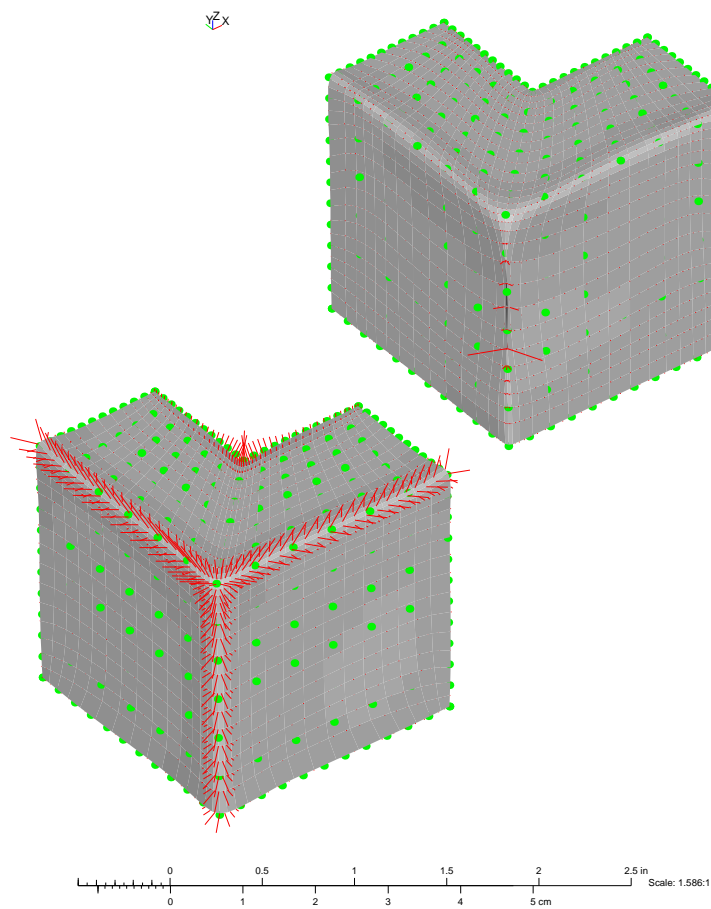


Abbildung 2.4: Approximation von Punkten auf einer Würfecke.
 Max. Fehler für die obere Fläche mit neuem Verfahren 0.40 mm,
 für die untere Fläche mit Verfahren [25] 0.36 mm

- (i) Das neue Approximationsverfahren erreicht maximale Lotfehler, die in der gleichen Größenordnung liegen wie beim Verfahren von [25]. Teilweise sind die Fehler in den präsentierten Beispielen sogar geringer als bei [25].
- (ii) Der Berechnungsaufwand des neuen Verfahrens zum Erreichen dieser Fehler ist im Vergleich zu [25] erheblich geringer. Besonders deutlich wird das anhand der Beispiele Punkte auf der Schraubfläche (Abbildung 2.2) und Punkte auf der Translationsfläche. Die Aussage wird unterstrichen durch die Grafik 2.5. Auf der x-Achse ist die Anzahl der de Boor-Punkte $(n+1)(r+1)$ angetragen, auf der y-Achse die benötigte Rechenzeit. Die Meßwerte kamen durch Approximation von 271 Punkte mit beiden Verfahren durch 2 Iterationen zustande ($k=p=4$). Das „+“ markiert die Zeiten für das neue Verfahren, der „*“ die Zeiten für das Verfahren von [25]. Aus den Meßwerten läßt sich ablesen, daß das Verhältnis der Rechenzeiten zwischen den beiden Verfahren um so größer ist, je kleiner das Verhältnis zwischen der Anzahl der de Boor-Punkte und der Anzahl der Vorgabepunkte ist. So ist das neue Verfahren bei 16 de Boor-Punkten ungefähr um den Faktor 80 schneller und bei 40 de Boor-Punkten ungefähr um den Faktor 8 schneller.
- (iii) Der belegte Speicherplatz spricht ebenfalls für das oben vorgestellte Verfahren. Das Verhältnis zur Speicherung der Koeffizienten der Gleichungssysteme vom neuen Verfahren zu [25] ist bei gleicher Anzahl der Segmente nämlich $16/(m+1)$ und somit bei einer größeren Anzahl von Punkten ($m > 100$), die bei Flächenapproximationen häufig auftritt, sehr klein.
- (iv) Das neue Verfahren unterliegt hinsichtlich der Ordnungen k, p prinzipiell keiner Beschränkung (siehe auch Kapitel Kurvenapproximation). Das Verfahren [25] kann dagegen nur bis zu Ordnungen $k, p < 5$ zufriedenstellend angewendet werden. Wie bei integralen B-Splines sollten im neuen Verfahren Ordnungen $k, p < 9$ gewählt werden, da sonst die Neigung zur Wellenbildung zunimmt.

2.3.2 Test zwischen rationaler und integraler Approximation

In den folgenden Beispielen werden das neue rationale Approximationsverfahren und das Verfahren mit integraler Approximation gegenübergestellt. Das Verfahren mit integraler Approximation minimiert

$$F_I := \sum_{i=0}^m (\mathbf{X}(u_i, v_i) - \mathbf{P}_i)^2 \quad (h_{jl} = 1) \quad (2.3)$$

Ein Beispiel enthält wieder die Darstellungen von zwei Flächen zu denselben Vorgabepunkten. Die obere Fläche wurde mit dem rationalen Approximationsverfahren, die untere mit dem integralen Approximationsverfahren bestimmt. Beide Flächen haben jeweils äquidistante Knotenvektoren \mathbf{U}, \mathbf{V} mit einfachen inneren Knoten. Die unten angeführten Fehlerwerte bezeichnen maximale Lotfehler zwischen den Flächen und den Vorgabepunkten. Zum Abschätzen der Verhältnisse zwischen Ausdehnung der Flächen und den Fehlerwerten ist in den Abbildungen ein Maßstab eingezeichnet. Wie in Abschnitt 2.3.1 sind an den Flächen

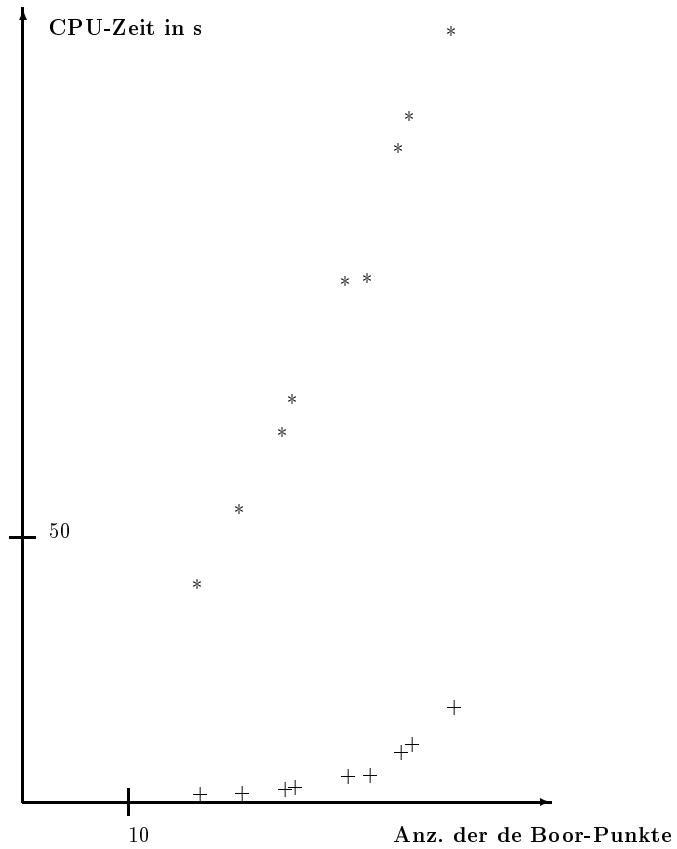


Abbildung 2.5: Rechenzeiten im Vergleich (* = Verfahren [25], + = neues Verfahren)

die Absolutkrümmungen in Richtung der Normalen als „Nadeln“ gezeichnet. Dabei ist die Nadellänge proportional zum Wert der Absolutkrümmung.

Beispiel 1 (Abbildung 2.6) zeigt die bikubischen Approximationsflächen von 1070 auf einem Pedaltopf abgetasteten Vorgabepunkten. Die obere rationale Fläche mit 8×8 Segmenten hat einen Fehler von 4.08 mm, die untere integrale Fläche mit 13×13 Segmenten einen Fehler von 4.95 mm. Die Rechenzeit für die obere Fläche belief sich auf 1 min 52.7 s bei 2 Iterationen und für die untere Fläche auf 2.4 s bei 1 Iteration. Bei beiden Flächen fällt der wellige Verlauf auf. Hauptgrund dafür ist, daß die Vorgabepunkte in Bereichen mit größeren Änderungen der Oberflächenform (z. B. in den Flankenbereichen) Lücken aufweisen.

Das zweite Beispiel stellt in Abbildung 2.7 626 Punkte, die auf einer Griffschale liegen, zusammen mit den Approximationsflächen dar. Die obere Fläche hat 2×8 Segmente und die untere 5×10 Segmente. Beide Flächen haben jeweils die Ordnungen 4×6 . Der Fehler von 2.27 mm der oberen Fläche wurde mit 3 Iterationen in 24.4 s, der Fehler von 2.10 mm der unteren Fläche mit 3 Iterationen in 3.4 s erreicht. Ähnlich wie in Beispiel 1 verlaufen die Flächen besonders in den Randbereichen wellig. Bei der integralen Fläche ist das wellige Verhalten erheblich ausgeprägter als bei der rationalen. Zudem weist die rationale Fläche eine „rundere“ Form auf.

Im folgenden Beispiel (Abbildung 2.8) sind 765 Punkte auf einem Seitenteil einer Autokarosserie vorgegeben, die von einer biquintischen rationalen Fläche mit 1×11 Segmenten bzw. einer biquintischen integralen Fläche mit 10×10 Segmenten approximiert werden. Der Fehler der oberen Fläche beträgt 0.16 mm mit einer Rechendauer von 3 min 26.7 s (9 Iterationen), der Fehler der unteren Fläche 0.32 mm mit einer Rechendauer von 8.8 s (10 Iterationen). Zur integralen Approximation ist anzumerken, daß sie auch mit mehr Segmenten den Fehler der rationalen Fläche nicht erreichen konnte.

Als viertes und letztes Beispiel enthält Abbildung 2.9 biquintische Approximationen von 181 Punkten auf einer Bézierfläche vom Grad 9×9 . Für die obere Fläche mit einem Segment bemißt sich der Fehler auf 0.021 mm, für die untere Fläche mit 2×2 Segmenten auf 0.018 mm. Die Rechenzeit für die obere Fläche (10 Iterationen) lag bei 11.9 s, die Rechenzeit für die untere Fläche (30 Iterationen) bei 3.7 s.

Aufgrund der Beispiele kann das neue rationale Approximationsverfahren wie folgt bewertet werden: Das neue Verfahren braucht, um ähnliche maximale Lotfehler zu erhalten, weniger Segmente, in den Beispielen 1-3 deutlich weniger Segmente, als das integrale Approximationsverfahren. Die dafür benötigte Rechenzeit ist allerdings wesentlich länger, kann aber in ihrer Größenordnung (maximal im Minutenbereich) noch toleriert werden. Der Unterschied in den Rechenzeiten erklärt sich durch die im Vergleich zum integralen Verfahren 4mal größere Koeffizientenmatrix, die beim Lösen des linearen Gleichungssystems auftritt, und daß beim integralen Verfahren außerdem die Bandstruktur der Koeffizientenmatrix ausgenutzt wird.

Um den welligen Verlauf der Approximationsflächen in den Beispielen 1 und 2 zu vermeiden, wird bei der integralen Approximation zum Beispiel (s. [4]) in (2.3) ein Energieterm hinzugefügt, der näherungsweise die Biegeenergie der dünnen Platte beschreibt. Statt (2.3) wird nun

$$\bar{F}_I := F_I + \lambda \iint \mathbf{X}_{uu}^2(u, v) + 2\mathbf{X}_{uv}^2(u, v) + \mathbf{X}_{vv}^2(u, v) du dv, \quad \lambda > 0 \quad (2.4)$$

mit einem fest gewählten Faktor λ minimiert, der den Einfluß des Energieterms steuert. Das

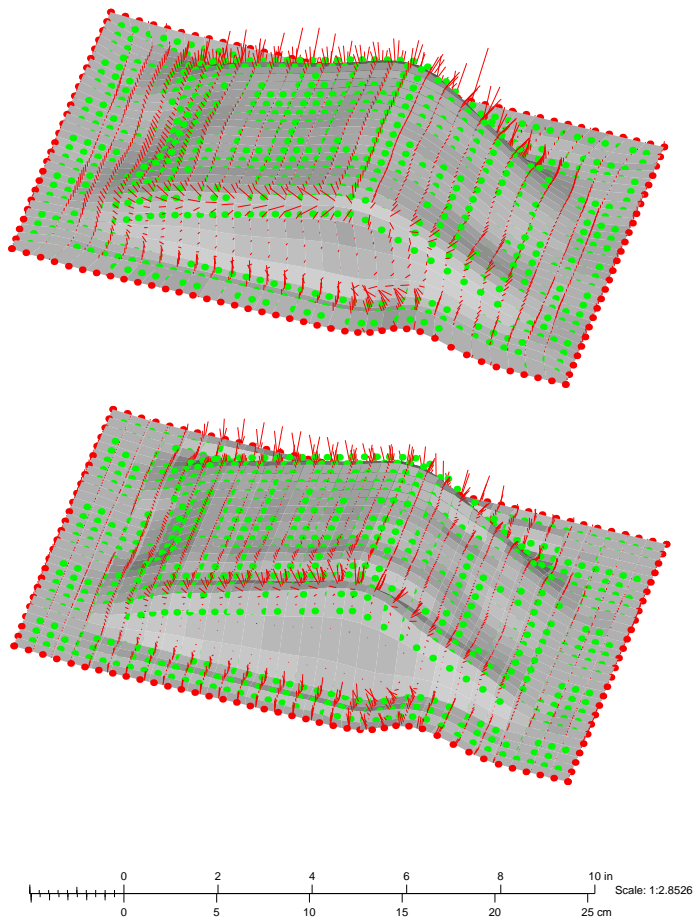


Abbildung 2.6: Approximation von Punkten zu Beispiel 1. Max. Fehler für die obere rationale Fläche 4.08 mm, für die untere integrale Fläche 4.95 mm

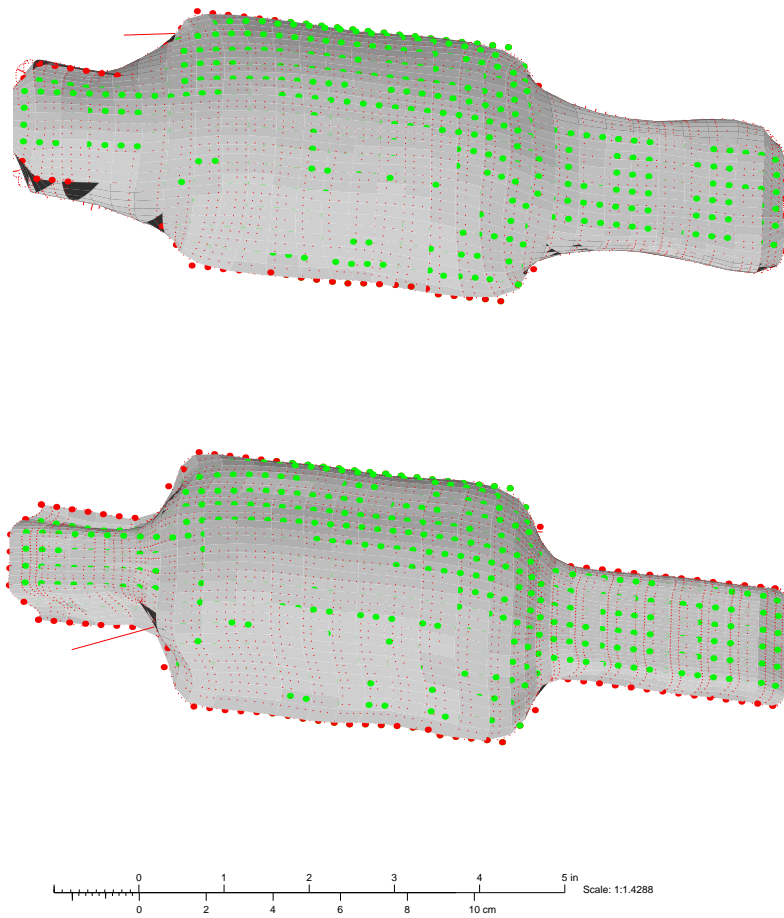


Abbildung 2.7: Approximation von Punkten zu Beispiel 2. Max. Fehler für die obere rationale Fläche 2.27 mm, für die untere integrale Fläche 2.10 mm

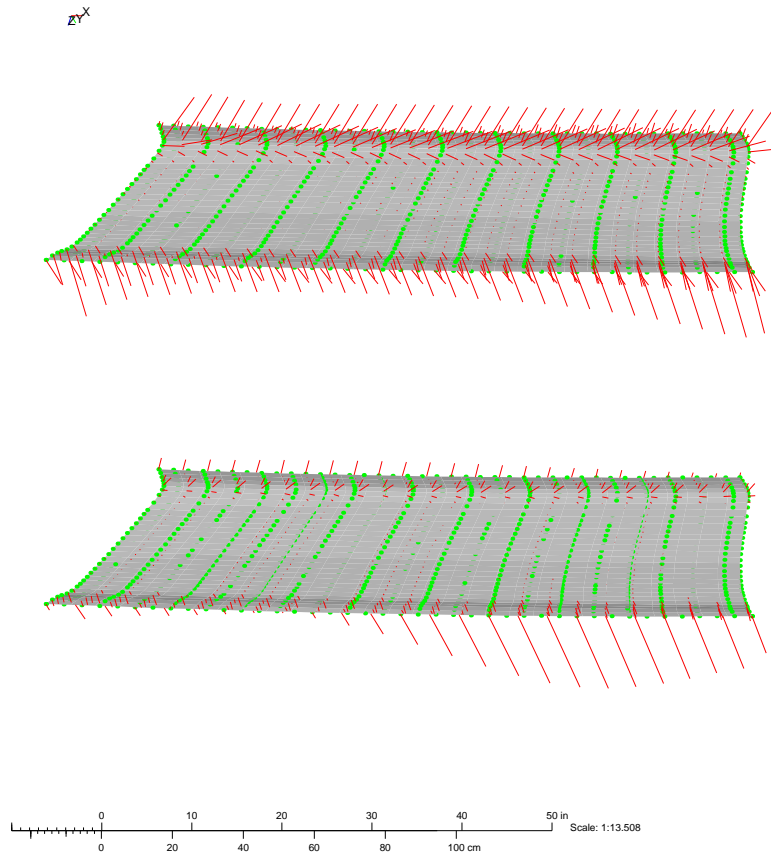


Abbildung 2.8: Approximation von Punkten zu Beispiel 3. Max. Fehler für die obere rationale Fläche 0.16 mm, für die untere integrale Fläche 0.32 mm

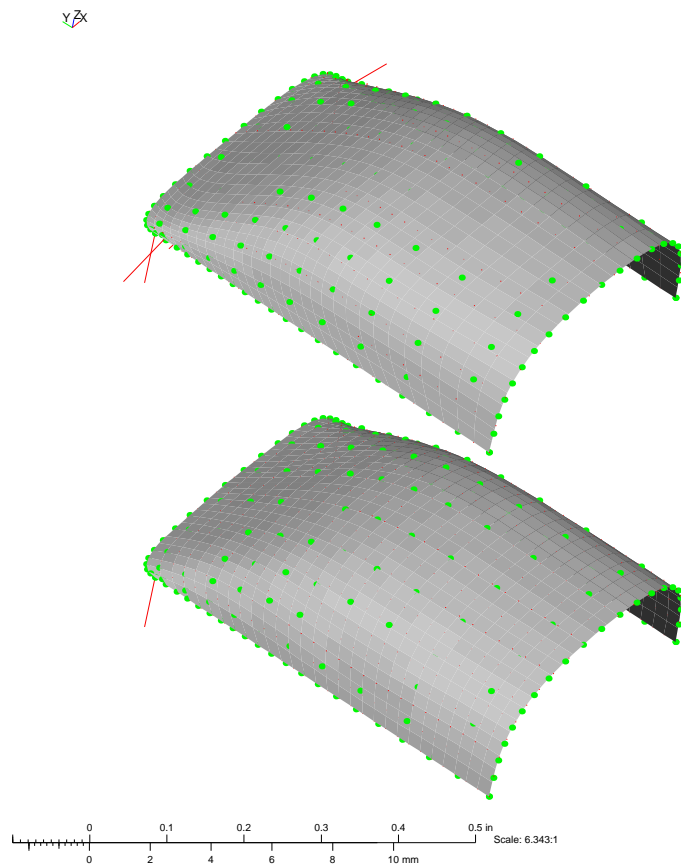


Abbildung 2.9: Approximation von Punkten zu Beispiel 4. Max. Fehler für die obere rationale Fläche 0.021 mm, für die untere integrale Fläche 0.018 mm

Ergebnis der Minimierung von (2.4) mit den Vorgabepunkten aus Beispiel 1 zeigt Abbildung 2.10. Verglichen mit den Flächen in Abbildung 2.6 ist die integrale Approximationsfläche (gleiche Segmentanzahl wie die integrale Fläche dort) stark geglättet bei ähnlichem maximalen Lotfehler. Ein analoges Vorgehen für diese Art von Vorgabepunkten bei dem neuen rationalen Verfahren wäre ebenfalls sinnvoll, bedarf aber noch umfangreicher Untersuchungen, die den Rahmen dieser Arbeit sprengen würde, wenn die Minimierung von \bar{F} plus Energieterm auf ein lineares Gleichungssystem führen soll.

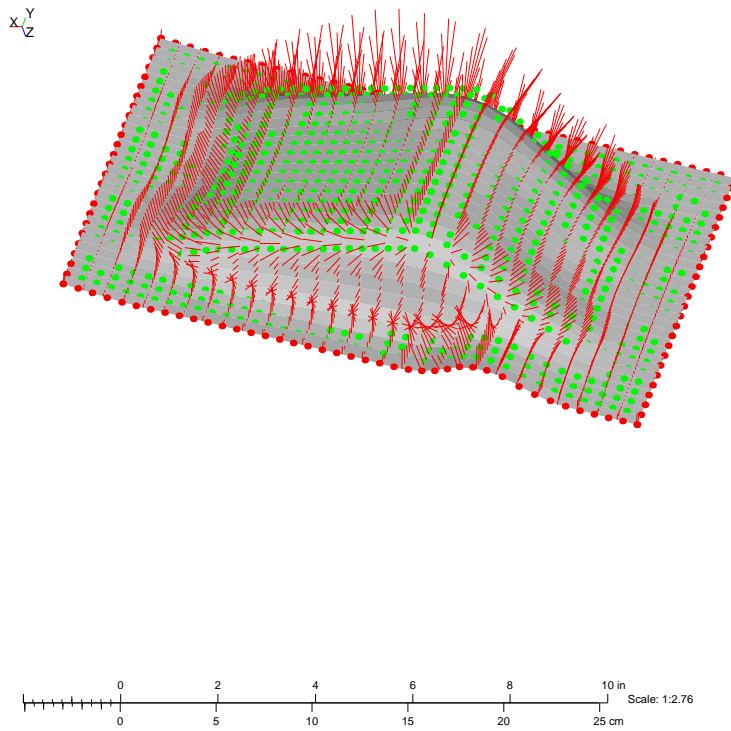


Abbildung 2.10: Approximation von Punkten aus Beispiel 1 mit integraler Fläche unter Hinzunahme eines Energieterms. Max. Fehler 4.97 mm

Kapitel 3

Segmentbestimmung bei Approximation mit B-Splinekurven

Bei der Approximation von gegebenen Punkten $\mathbf{P}_i \in \mathbb{R}^3$ und den Parameterwerten t_i ($i = 0, \dots, m$) mit einer rationalen oder integralen ($h_j = 1$) B-Splinekurve

$$\mathbf{X}(t) := \frac{\sum_{j=0}^n h_j \mathbf{d}_j N_{jk}(t)}{\sum_{j=0}^n h_j N_{jk}(t)} \quad , \quad \mathbf{d}_j \in \mathbb{R}^3 \quad , \quad h_j \in \mathbb{R}$$

der Ordnung k definiert über dem Knotenvektor

$$\mathbf{T} := (\tau_0 = \dots = \tau_{k-1}, \tau_k, \dots, \tau_n, \tau_{n+1} = \dots = \tau_{n+k}) \quad , \quad \tau_{k-1} < \tau_k < \dots < \tau_{n+1} \quad , \quad n \geq k-1$$

soll die Approximationskurve $\mathbf{X}(t)$ meist die folgenden Bedingungen erfüllen:

- sie soll Anfangs- und Endpunkt interpolieren

$$\mathbf{X}(t_0) = \mathbf{X}(\tau_{k-1}) = \mathbf{P}_0 \quad , \quad \mathbf{X}(t_m) = \mathbf{X}(\tau_{n+1}) = \mathbf{P}_m$$

- und eine vorgegebene Fehlertoleranz ε einhalten

$$\sum_{i=0}^m (\mathbf{X}(t'_i) - \mathbf{P}_i)^2 \leq \varepsilon \quad (3.1)$$

oder alternativ

$$\max_i |\mathbf{X}(t'_i) - \mathbf{P}_i| \leq \varepsilon$$

Dabei sind die t'_i so gewählt, daß die Fehlervektoren $(\mathbf{X}(t'_i) - \mathbf{P}_i)$ ungefähr senkrecht auf den Tangentenvektoren $\dot{\mathbf{X}}(t'_i)$ ($i = 1, \dots, m-1$) stehen.

(3.1) stellt eine Bedingung an die Segmentanzahl $s = n - k + 2$ der Approximationskurve. Falls s zu klein ist, kann (3.1) nicht eingehalten werden.

Für integrale B-Splinekurven ($h_j = 1$) sind im CAGD schon mehrere Verfahren zur Lösung von (3.1) mit möglichst kleiner Segmentanzahl entwickelt worden ([3], [5], [15], [25]). Hierbei handelt es sich hauptsächlich um Knoteneliminationsverfahren ([15], [5]): Zuerst werden die gegebenen Punkte mit entsprechend vielen Segmenten interpoliert oder so approximiert, daß die Fehlertoleranz ε deutlich unterschritten wird. Dann werden in einem Iterationsverfahren pro Schritt ein oder mehrere weniger wirksame Knoten entfernt. Die vorhandenen Knoten werden in jedem Iterationsschritt bei einer gegebenen Toleranz gewichtet und im Fall der Entfernung mehrerer Knoten wird mit einem Suchverfahren ermittelt, wieviele Knoten bei Einhaltung der Toleranz gemäß der Wichtungsreihenfolge entfernt werden können. Die optimale Approximationskurve ist gefunden, wenn kein Knoten mehr eliminiert werden kann.

In [25] wird ein Verfahren zur Bestimmung der Segmentanzahl für integrale kubische ($k = 4$) B-Splinekurven beschrieben. Vor Beginn des Approximationsprozesses wird der entstehende Fehler durch Übertragung funktionswertiger Fehlerabschätzungen aus der Spline Analysis auf parametrisierte Kurven und Anwendung dieser Ergebnisse auf das Problem (3.1) abgeschätzt. Aus dem Schätzwert für den Fehler und der vorgegebenen Fehlertoleranz ε wird ein Vorschlagswert für die Segmentanzahl bestimmt.

Hier wird nun ein Verfahren vorgestellt, daß für B-Splinekurven (rational oder integral) eine möglichst kleine Segmentanzahl unter Einhaltung von (3.1) bestimmt.

3.1 Iterative Segmentbestimmung

Das im folgenden beschriebene Verfahren verbessert eine in [3] entwickelte Methode. Die Idee des Verfahrens ist es, eine Stelle \bar{s} der diskreten Funktion $f(s) := F_s - \varepsilon$ zu finden, für die $f(\bar{s}) \leq 0$ und \bar{s} möglichst klein ist. $F_s := \sum_{i=0}^m (\mathbf{X}_s(t_i) - \mathbf{P}_i)^2$ bezeichnet die Summe der Fehlerquadrate zwischen der Approximationskurve \mathbf{X}_s mit s Segmenten und den Punkten \mathbf{P}_i . Die Approximationskurve wird mit dem Verfahren aus Abschnitt 1.3 berechnet. Auf die Anordnung der Knoten im Knotenvektor \mathbf{T} wird später näher eingegangen.

Bei kontinuierlichen Funktionen entspricht diese Idee der Bestimmung von Nullstellen. Dazu existieren zahlreiche Verfahren (z. B. Intervallhalbierung, regula falsi und für außerdem stetig differenzierbare Funktionen das Newton-Verfahren). Eines dieser Verfahren zur Nullstellenbestimmung ist das *Sekantenverfahren*, das eine mehr als lineare Konvergenzordnung hat und keine Ableitung der Funktion benötigt. Das Sekantenverfahren kann daher auch in leicht abgewandelter Form auf die diskrete Funktion f angewendet werden, um eine möglichst kleine Segmentanzahl \bar{s} mit $f(\bar{s}) \leq 0$ zu erhalten: Bevor das Iterationsverfahren dargestellt wird, werden einige Bezeichnungen erläutert. s_j ($j = 0, 1, \dots$) bezeichnet die Anzahl der Segmente beim j -ten Iterationsschritt. $\lceil x \rceil$ ist die ganze Zahl, für die gilt $x - 1 < \lceil x \rceil \leq x$. Die Iteration ist damit zunächst wie folgt definiert:

$$\begin{aligned} & \text{Startwert } s_0 \\ s_{j+1} &= s_j + \Delta s_j \quad (j = 0, 1, \dots) \\ \Delta s_j &:= \begin{cases} \text{sign}(f(s_0)), & j = 0 \\ \text{sign}(f(s_j)) \cdot \Delta_1, & j = 1, 2, \dots \end{cases} \end{aligned}$$

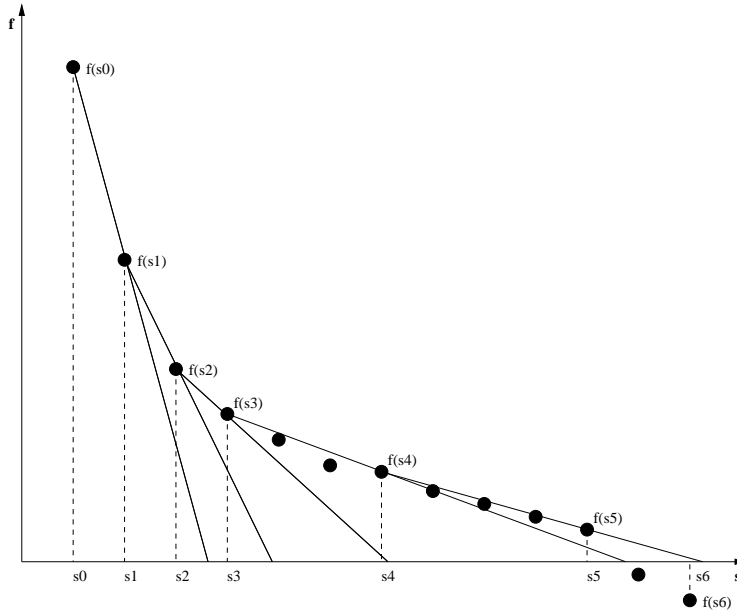


Abbildung 3.1: Geometrische Interpretation der Iteration (3.2)

$$\Delta_1 := \left\lceil \left| \frac{f(s_j)}{f(s_{j-1}) - f(s_j)} \Delta_{s_{j-1}} \right| \right\rceil \quad (3.2)$$

$$\text{sign}(x) := \begin{cases} 1 & , \quad x > 0 \\ -1 & , \quad \text{sonst} \end{cases}$$

Die Iteration (3.2) wird solange durchgeführt bis $f(s_j) \leq 0$ und $\Delta_{s_{j-1}} > 0$ sind. Im Unterschied zu [3] kann bei der Iteration (3.2) die Segmentanzahl s_j erhöht oder erniedrigt werden. Somit ist man nicht (wie in [3]) gezwungen, den Startwert s_0 für die Segmentanzahl gleich Eins zu wählen, um eine möglichst kleine Segmentanzahl zu erhalten, falls kein Schätzwert für s_0 vorliegt. Abbildung 3.1 veranschaulicht geometrisch den Ablauf der Iteration (3.2). Dort sind schwarz gepunktet die Funktionswerte $f(s)$ für die entsprechenden Segmentanzahlen s angetragen. Die Funktionsargumente s zweier benachbarter Funktionswerte $f(s)$ unterscheiden sich jeweils um ein Segment. Die Iteration bricht hier bei der Segmentanzahl s_6 ab.

Wie zum Beispiel in Abbildung 3.1 erkennbar ist, wo $\Delta_{s_2} = 1$, aber $\Delta_{s_3} = 3$ und $\Delta_{s_4} = 4$ ist, führt die Anwendung der Iteration (3.2) jedoch oft zu Situationen, in denen die Schrittweite Δ_{s_j} abwechselnd sehr groß und dann wieder sehr klein ist. Sinnvoller wäre

es, eine allmähliche Veränderung der Schrittweite zu haben. Daher führen wir obere und untere Schranken für die Schrittweite ein, um sie zu dämpfen. Eine untere Schranke Δ_2 und eine obere Schranke Δ_3 kann aus der Schrittweite des vorhergehenden Iterationsschritts festgelegt werden:

$$\begin{aligned}\Delta_2 &:= \lceil |\Delta s_{j-1}/2| \rceil \\ \Delta_3 &:= 2|\Delta s_{j-1}|\end{aligned}$$

Beide Schranken sind Erfahrungswerte, die in den Beispielen zu guten Resultaten führten. Eine weitere obere Schranke Δ_4 der Schrittweite ergibt sich aus der Forderung, daß die Anzahl der Segmente und damit die Anzahl der Unbekannten (Anzahl der Komponenten von \mathbf{d}_j und Anzahl der h_j) kleiner ist als die Anzahl der Punktkoordinaten der \mathbf{P}_i . Anderenfalls können im linearen System zur Bestimmung der Approximationskurve Lücken entstehen.

$$\Delta_4 := \left\lceil \frac{3}{4}m - k + 2 - s_j \right\rceil$$

Schließlich ist 1 eine triviale untere Schranke der Schrittweite. Eine geeignetere Schrittweite als in (3.2) bestimmt sich nun, indem zuerst das Maximum der unteren Schranken 1, Δ_2 zusammen mit Δ_1 bestimmt wird und dann das Minimum der oberen Schranken Δ_3 , Δ_4 mit diesem Maximum. Außerdem muß das Vorzeichen von $f(s_j)$ berücksichtigt werden, um festzustellen, ob die Segmentanzahl erhöht oder vermindert werden soll. Wir erhalten somit ein modifiziertes Iterationsverfahren, wenn Δs_j in (3.2) durch

$$\Delta s_j := \begin{cases} \text{sign}(f(s_0)) , & j = 0 \\ \text{sign}(f(s_j)) \cdot \min\{\max\{1, \Delta_1, \Delta_2\}, \Delta_3, \Delta_4\} , & j = 1, 2, \dots \end{cases} \quad (3.3)$$

ersetzt wird.

(3.3) findet, vorausgesetzt daß eine Segmentanzahl s^* existiert, so daß alle Segmentanzahlen s mit $s \geq s^*$ (3.1) erfüllen, in Abhängigkeit vom Startwert s_0 eine Segmentanzahl s_j , die (3.1) einhält. Denn ist $f(s_0) > 0$, dann erhält man eine Iterationsfolge $s_0 < s_1 < \dots < s_j$ mit $f(s_1), \dots, f(s_{j-1}) > 0$ und $f(s_j) \leq 0$. Ist dagegen $f(s_0) \leq 0$, so ergibt sich die Folge $s_0 > s_1 > \dots > s_k < s_{k+1} < \dots < s_j$ mit $f(s_1), \dots, f(s_{k-1}) \leq 0$, $f(s_k), \dots, f(s_{j-1}) > 0$ und $f(s_j) \leq 0$.

Nun gehen wir noch auf die Anordnung der Knoten bei fester Segmentanzahl $s_j = n_j - k + 2$ im Knotenvektor \mathbf{T} ein. Dazu wird zuerst eine Approximation mit dem äquidistanten Knotenvektor

$$\mathbf{T} = (\underbrace{0, \dots, 0}_{k-fach}, \frac{1}{n_j - k + 2}, \dots, \frac{n_j - k + 1}{n_j - k + 2}, \underbrace{1, \dots, 1}_{k-fach})$$

durchgeführt, in der eine B-Splinekurve \mathbf{X}_{s_j} mit Parameterwerten t'_i und die Fehlerquadratsumme F_1 bestimmt wird. Anschließend wird mit einem nichtäquidistanten Knotenvektor \mathbf{T}' nochmal approximiert und die Fehlerquadratsumme F_2 ausgerechnet. Die Knoten von $\mathbf{T}' = (0, \dots, 0, \tau_k, \dots, \tau_n, 1, \dots, 1)$ werden so bestimmt, daß die aufsummierten Fehlerquadrate $(\mathbf{X}_{s_j}(t'_i) - \mathbf{P}_i)^2$ der Approximationskurve mit äquidistantem Knotenvektor und der

Punkte zwischen zwei benachbarten Knoten τ_l und τ_{l+1} ungefähr der mittleren Fehlerquadratsumme F_1 dividiert durch die Anzahl der Segmente s_j entsprechen oder formal

$$\sum_{\tau_l \leq t'_i < \tau_{l+1}} (\mathbf{X}_{s_j}(t'_i) - \mathbf{P}_i)^2 \approx \frac{F_1}{s_j} \quad (l = k-1, \dots, n) \quad (3.4)$$

(3.4) liegt das Ziel zugrunde, daß in den Bereichen des Knotenvektors \mathbf{T}' , in denen die Approximationsfehler mit dem äquidistanten Knotenvektor \mathbf{T} größer waren, mehr Knoten liegen sollen als in Bereichen, in denen die Approximationsfehler mit \mathbf{T} kleiner waren. $f(s_j)$ wird dann gesetzt als $f(s_j) := F_{s_j} - \varepsilon$ mit $F_{s_j} := \min\{F_1, F_2\}$. Das Iterationsverfahren (3.3) kann alternativ auch auf die Funktion $f(s) := \bar{F}_s - \varepsilon$ mit $\bar{F}_s := \max_i |\mathbf{X}(t'_i) - \mathbf{P}_i|$ angewendet werden.

Bei Durchführung der Iteration (3.3) im kubischen Fall kann als guter Startwert die mit Methode [25] geschätzte Segmentanzahl

$$s_0 = \left\lceil \frac{\sqrt{2}}{\pi} \frac{\|\mathbf{S}_8^{(4)}\|_2^{\frac{1}{4}}}{\varepsilon^{\frac{1}{8}}} \right\rceil + 1 \quad (3.5)$$

verwendet werden. In (3.5) sind $\|\mathbf{X}\|_2^2 := \int_0^1 |\mathbf{X}(t)|^2 dt$ und \mathbf{S}_8 ist die integrale B-Splinekurve der Ordnung 8 mit äquidistantem Knotenvektor, die die Bedingungen

$$\begin{aligned} \mathbf{S}_8 &= \mathbf{P}_i \quad (i = 0, \dots, m) \\ \mathbf{S}_8^{(\lambda)}(t_0) &= \mathbf{S}_8^{(\lambda)}(t_m) = 0 \quad (\lambda = 4, 5, 6) \end{aligned}$$

erfüllt.

3.2 Ergebnisse

Die Arbeitsweise des vorgestellten Iterationsverfahrens soll durch einige Beispiele veranschaulicht werden. Hierbei wird in jedem Beispiel der Iteration der Segmentanzahl mit rationalen B-Splinekurven auch die Iteration gegenübergestellt, in der in jedem Schritt mit integralen B-Splinekurven ($h_j = 1$) approximiert wird. Bei der integralen bzw. rationalen Approximation mit äquidistantem und mit nichtäquidistantem Knotenvektor in jedem Iterationsschritt wurden jeweils maximal 20 Parameterkorrekturschritte durchgeführt.

Im ersten Beispiel sind 101 Punkte auf einer Schraubenlinie gegeben. Sie sollen mit einer quintischen ($k = 6$) B-Splinekurve, einer Fehlertoleranz $\varepsilon = 0.05 \text{ mm}^2$ und der Startsegmentanzahl $s_0 = 13$ approximiert werden. Tabelle 3.1 zeigt die Resultate des Verfahrens mit rationaler Approximation und Abbildung 3.2 die rationale Approximationskurve, bei der das Verfahren abbricht. Entsprechendes zeigen Tabelle 3.2 und Abbildung 3.3 für das Verfahren mit integraler Approximation. In der ersten Zeile der Tabelle sind ausgehend von der Startsegmentanzahl s_0 der Reihenfolge nach alle von der Iteration ermittelten Segmentanzahlen aufgeführt. In den Zeilen 2 und 3 stehen die zugehörigen Fehlerquadratsummen F_1 und F_2 . Man sieht, daß bei wachsender Segmentanzahl die Fehlerquadratsumme F_{s_j} monoton fällt. Des weiteren treten Schrittweiten Δs_j mit $|\Delta s_j| \geq 1$ auf. $|\Delta s_j|$ ist groß, wenn die relative

Segmentanz.	$s_0 = 13$	$s_1 = 12$	$s_2 = 10$	$s_3 = 6$	$s_4 = 1$
F_1 (mm ²)	0.019	0.02	0.026	0.034	383
F_2 (mm ²)	0.028	0.12	0.034	0.030	383

Segmentanz.	$s_5 = 3$	$s_6 = 4$	$s_7 = 5$	$s_8 = 6$
F_1 (mm ²)	13.3	0.75	0.10	0.034
F_2 (mm ²)	7.19	0.36	0.067	0.030

Tabelle 3.1: Iteration der Segmentanzahl mit rationalen B-Splines bei schraubförmigen Punkten und $\varepsilon = 0.05$ mm²

Segmentanz.	$s_0 = 13$	$s_1 = 12$	$s_2 = 10$	$s_3 = 6$	$s_4 = 8$	$s_5 = 9$
F_1 (mm ²)	0.019	0.021	0.023	0.53	0.072	0.034
F_2 (mm ²)	0.11	0.062	0.087	0.22	0.21	0.049

Tabelle 3.2: Iteration der Segmentanzahl mit integralen B-Splines bei schraubförmigen Punkten und $\varepsilon = 0.05$ mm²

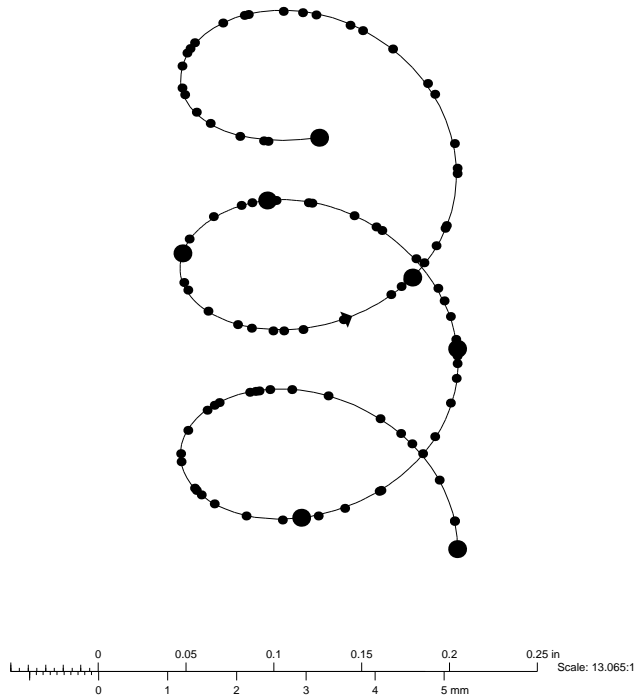
Änderung zwischen $F_{s_{j-1}}$ und F_{s_j} klein ist (z. B. $|\Delta s_1| = |s_2 - s_1| = 2$, $|\Delta s_2| = 4$, $|\Delta s_3| = 5$). $|\Delta s_j|$ ist klein, wenn die relative Änderung zwischen $F_{s_{j-1}}$ und F_{s_j} groß ist (z. B. $|\Delta s_5| = 1$, $|\Delta s_6| = 1$). Bei den Fehlerquadratsummen F_1 mit äquidistantem Knotenvektor und F_2 mit nichtäquidistantem Knotenvektor sind keine wesentlichen Unterschiede zu erkennen. Auch bei integraler Approximation fällt die Fehlerquadratsumme F_{s_j} monoton bei wachsender Segmentanzahl und $|\Delta s_j|$ ist groß, wenn die relative Änderung zwischen $F_{s_{j-1}}$ und F_{s_j} klein ist (z. B. $|\Delta s_1| = 2$, $|\Delta s_2| = 4$). Umgekehrt ist auch $|\Delta s_j|$ klein, wenn die relative Änderung zwischen $F_{s_{j-1}}$ und F_{s_j} groß ist (z. B. $|\Delta s_4| = 1$). Dagegen bricht hier das Verfahren erst bei 9 Segmenten ab, während es mit rationaler Approximation schon bei 6 Segmenten abbricht. Zudem sind hier die Fehlerquadratsummen F_1 bei den Segmentanzahlen s_0, s_1, s_2 wesentlich geringer als die Fehlerquadratsummen F_2 .

Tabellen 3.3 und 3.4 zeigen zu den gleichen Vorgabepunkten und Fehlertoleranz wie in Beispiel 1 die Iterationsergebnisse bei der Startsegmentanzahl $s_0 = 2$. Da $f(s_0) > 0$ im integralen wie im rationalen Fall ist, ergeben sich monoton steigende Folgen von Segmentanzahlen.

Das nächste Beispiel (Tabellen 3.5, 3.6, 3.7, 3.8) enthält die Iterationsdaten zu den Vorgabepunkten aus Beispiel 1 und der Fehlertoleranz $\varepsilon = 0.5$ mm². Abbildung 3.4 und 3.5 zeigen die Approximationskurven, die (3.1) einhalten. Die beiden Iterationsfolgen mit rationaler Approximation und die Iterationsfolge mit integraler Approximation und Startsegmentanzahl $s_0 = 2$ entsprechen denen aus den vorangegangenen Beispielen mit dem Unterschied,

Segmentanz.	$s_0 = 2$	$s_1 = 3$	$s_2 = 4$	$s_3 = 5$	$s_4 = 6$
F_1 (mm ²)	142	13.3	0.75	0.10	0.034
F_2 (mm ²)	102	7.19	0.36	0.067	0.030

Tabelle 3.3: Iteration der Segmentanzahl mit rationalen B-Splines bei schraubförmigen Punkten und $\varepsilon = 0.05$ mm²



$$\mathbf{T}' = (0, \dots, 0, 0.259, 0.364, 0.483, 0.545, 0.684, 1, \dots, 1)$$

Abbildung 3.2: Approximation von schraubförmig gelegenen Punkten. Nichtäquidistante rat. B-Splinekurve mit 6 Segmenten, Fehlerquadratsumme 0.030 mm^2 und max. Fehler 0.023 mm (fett gezeichnete Punkte sind die den Knoten entsprechenden Kurvenpunkte)

Segmentanz.	$s_0 = 2$	$s_1 = 3$	$s_2 = 4$	$s_3 = 5$	$s_4 = 6$	$s_5 = 7$	$s_6 = 9$
$F_1 \text{ (mm}^2\text{)}$	445	174	7.94	3.39	0.53	0.21	0.034
$F_2 \text{ (mm}^2\text{)}$	450	210	23.0	94.7	0.22	0.17	0.049

Tabelle 3.4: Iteration der Segmentanzahl mit integralen B-Splines bei schraubförmigen Punkten und $\varepsilon = 0.05 \text{ mm}^2$

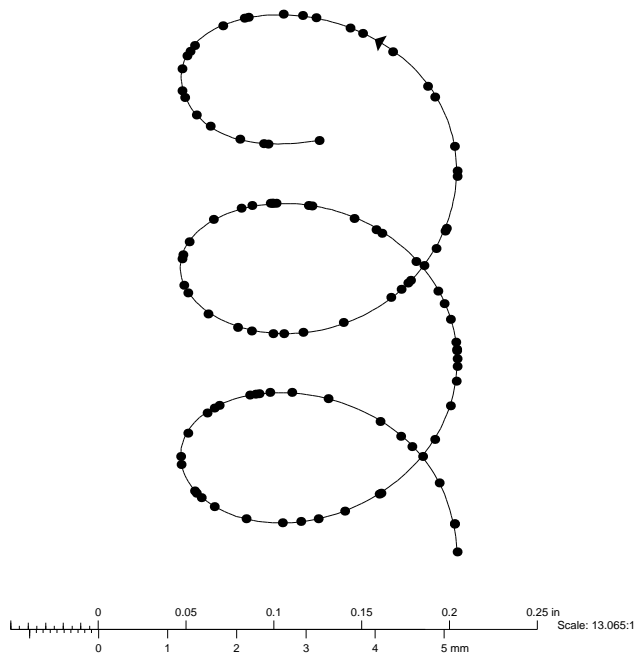


Abbildung 3.3: Approximation von schraubförmig gelegenen Punkten. Äquidistante int. B-Splinekurve mit 9 Segmenten, Fehlerquadratsumme 0.034 mm^2 und max. Fehler 0.027 mm

Segmentanz.	$s_0 = 13$	$s_1 = 12$	$s_2 = 10$	$s_3 = 6$	$s_4 = 1$	$s_5 = 3$	$s_6 = 4$
F_1 (mm ²)	0.019	0.02	0.026	0.034	383	13.3	0.75
F_2 (mm ²)	0.028	0.12	0.034	0.030	383	7.19	0.36

Tabelle 3.5: Iteration der Segmentanzahl mit rationalen B-Splines bei schraubförmigen Punkten und $\varepsilon = 0.5$ mm²

Segmentanz.	$s_0 = 13$	$s_1 = 12$	$s_2 = 10$	$s_3 = 6$	$s_4 = 1$
F_1 (mm ²)	0.019	0.021	0.023	0.53	467
F_2 (mm ²)	0.11	0.062	0.087	0.22	467

Segmentanz.	$s_5 = 3$	$s_6 = 4$	$s_7 = 5$	$s_8 = 6$
F_1 (mm ²)	174	7.94	3.39	0.53
F_2 (mm ²)	210	23.0	94.7	0.22

Tabelle 3.6: Iteration der Segmentanzahl mit integralen B-Splines bei schraubförmigen Punkten und $\varepsilon = 0.5$ mm²

daß hier wegen der größeren Fehlertoleranz schon bei 4 bzw. 6 Segmenten abgebrochen wird. Der Grund für die Übereinstimmung der jeweiligen Iterationsfolgen ist, daß bei ihnen s_0 und Δs_0 gleich sind und im weiteren dann immer die gleichen unteren Schranken Δ_2 , 1 oder die obere Schranke Δ_3 bei der Berechnung von Δs_j in (3.3) zum Tragen kommen. Im Gegensatz dazu wird bei der Iterationsfolge mit integraler Approximation und Startsegmentanzahl $s_0 = 13$ die Segmentanzahl $s_3 = 6$ auf $s_4 = 1$ verringert, während sie bei der Iterationsfolge mit Fehlertoleranz $\varepsilon = 0.05$ mm² (Tabelle 3.2) von $s_3 = 6$ auf $s_4 = 8$ erhöht wird. Ursache dafür ist, daß bei $\varepsilon = 0.05$ mm² $f(s_3) > 0$ und bei $\varepsilon = 0.5$ mm² $f(s_3) < 0$ ist. Wiederum führt das Iterationsverfahren mit rationaler Approximation zu einer Kurve mit weniger Segmenten ($s = 4$) als das Iterationsverfahren mit integraler Approximation ($s = 6$).

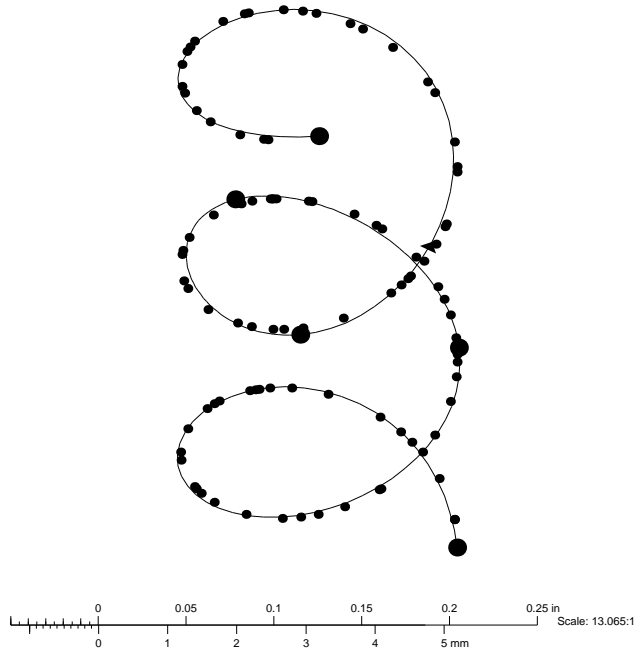
In den Tabellen 3.9 und 3.10 sind die Iterationsfolgen zu 41 M-förmig vorgegebenen Punkten mit kubischen ($k = 4$) B-Splinekurven und der Fehlertoleranz $\varepsilon = 0.5$ mm² zu sehen. Die Abbildungen 3.6 und 3.7 stellen die am Iterationsende resultierenden Approximationskurven dar. Die Folge der Fehlerquadratsummen F_{s_j} mit rationaler Approximation ist monoton fallend. Auch hier ist z. B. $\Delta s_3 = 4$ groß, da die relative Änderung zwischen F_{s_2} und F_{s_3} gering ist. Hingegen ist die Folge der F_{s_j} mit integraler Approximation nicht monoton fallend ($F_{s_2} > F_{s_1}$, $F_{s_4} > F_{s_3}$). Dennoch erhöht das Verfahren die Segmentanzahl weiter, da $f(s_j) > 0$ ($j = 0, \dots, 5$). Abweichend von den bisherigen Beispielen ist die Endsegmentanzahl bei der Iterationsfolge mit rationaler Approximation ($s_5 = 11$) größer als die

Segmentanz.	$s_0 = 2$	$s_1 = 3$	$s_2 = 4$
F_1 (mm ²)	142	13.3	0.75
F_2 (mm ²)	102	7.19	0.36

Tabelle 3.7: Iteration der Segmentanzahl mit rationalen B-Splines bei schraubförmigen Punkten und $\varepsilon = 0.5$ mm²

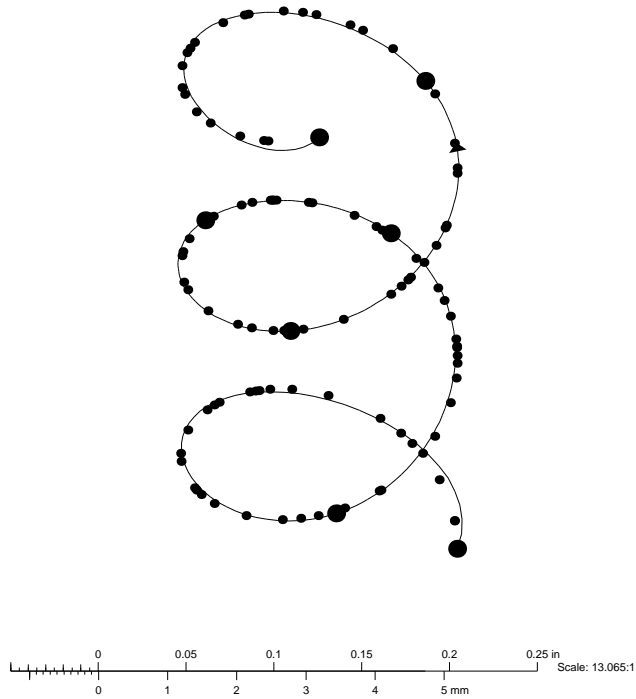
Segmentanz.	$s_0 = 2$	$s_1 = 3$	$s_2 = 4$	$s_3 = 5$	$s_4 = 6$
F_1 (mm ²)	445	174	7.94	3.39	0.53
F_2 (mm ²)	450	210	23.0	94.7	0.22

Tabelle 3.8: Iteration der Segmentanzahl mit integralen B-Splines bei schraubförmigen Punkten und $\varepsilon = 0.5$ mm²



$$\mathbf{T}' = (0, \dots, 0, 0.371, 0.494, 0.628, 1, \dots, 1)$$

Abbildung 3.4: Approximation von schraubförmig gelegenen Punkten. Nichtäquidistante rat. B-Splinekurve mit 4 Segmenten, Fehlerquadratsumme 0.36 mm² und max. Fehler 0.11 mm (fett gezeichnete Punkte sind die den Knoten entsprechenden Kurvenpunkte)



$$\mathbf{T}' = (0, \dots, 0, 0.273, 0.425, 0.511, 0.631, 0.779, 1, \dots, 1)$$

Abbildung 3.5: Approximation von schraubförmig gelegenen Punkten. Nichtäquidistante int. B-Splinekurve mit 6 Segmenten, Fehlerquadratsumme 0.22 mm^2 und max. Fehler 0.13 mm (fett gezeichnete Punkte sind die den Knoten entsprechenden Kurvenpunkte)

Segmentanz.	$s_0 = 1$	$s_1 = 2$	$s_2 = 3$	$s_3 = 5$	$s_4 = 9$	$s_5 = 11$
F_1 (mm ²)	14900	61.3	64.5	45.1	3.64	0.71
F_2 (mm ²)	14900	61.3	55.1	38.7	7.42	0.047

Tabelle 3.9: Iteration der Segmentanzahl mit rationalen B-Splines bei M-förmigen Punkten und $\varepsilon = 0.5$ mm²

Segmentanz.	$s_0 = 1$	$s_1 = 2$	$s_2 = 3$	$s_3 = 4$	$s_4 = 5$	$s_5 = 6$	$s_6 = 8$
F_1 (mm ²)	6770	1590	2230	28.1	29.8	25.5	10.2
F_2 (mm ²)	6770	1590	2590	18.5	30.3	27.9	0.42

Tabelle 3.10: Iteration der Segmentanzahl mit integralen B-Splines bei M-förmigen Punkten und $\varepsilon = 0.5$ mm²

der Iterationsfolge mit integraler Approximation ($s_6 = 8$). Dies hat seine Ursache im Iterationsverfahren. Denn aus Tabelle 3.11 des nächsten Beispiels entnimmt man, daß die rationale Approximation mit 7 Segmenten nur eine Fehlerquadratsumme von 0.07 mm² hat (im Vergleich zu 0.42 mm² der integralen Approximation mit 8 Segmenten). Diese Segmentanzahl ist aber in der Iterationsfolge nicht enthalten.

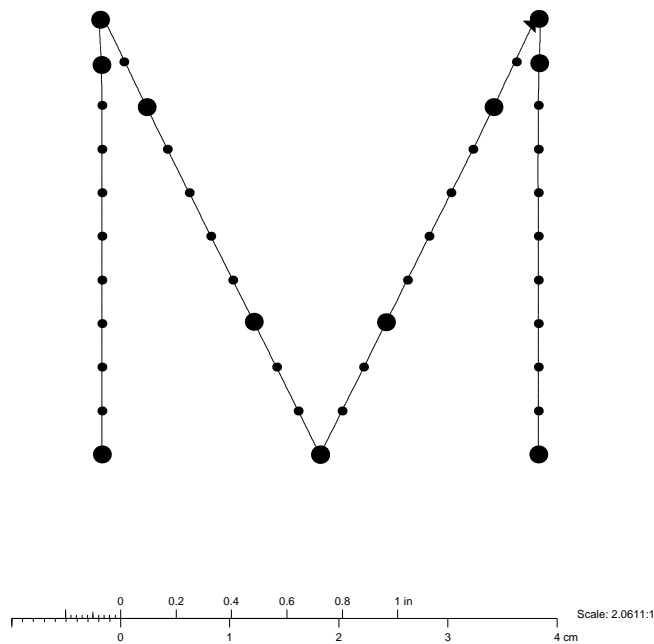
Das letzte Beispiel (Tabellen 3.11 und 3.12) zeigt zu gleichen Vorgabepunkten wie im vorhergehenden Beispiel und der Fehlertoleranz $\varepsilon = 5.0$ mm² die Ergebnisse der Iterationen. Anders als in den übrigen Beispielen ist die Folge der Fehlerquadratsummen F_{s_j} mit rationaler Approximation bei kleiner werdender Segmentanzahl nicht monoton wachsend (z. B. $F_{s_1} > F_{s_2}$, $F_{s_3} > F_{s_4}$). Trotzdem stellt das Verfahren sicher, daß die Segmentanzahl weiter verringert wird. Es gilt nämlich $f(s_j) < 0$ ($j = 0, \dots, 5$). Weiter fällt auf, daß bei den Segmentanzahlen s_0, s_3, s_5 die Fehlerquadratsumme F_2 deutlich kleiner als F_1 ist. Bei der Iteration mit integraler Approximation ist zu bemerken, daß sie bei 11 Segmenten abbricht, wohingegen sie im vorhergehenden Beispiel mit Fehlertoleranz $\varepsilon = 0.5$ mm² bei 8 Segmenten endet. Daran zeigt sich, daß die sich beim Iterationsverfahren ergebende Segmentanzahl von der Startsegmentanzahl s_0 abhängt.

Die Ergebnisse der vorangehenden Beispiele unterstreichen, daß das Iterationsverfahren sowohl für integrale als auch für rationale Kurvenapproximationen günstige also recht kleine Segmentanzahlen bestimmt, bei denen die Fehlertoleranz ε eingehalten wird.

Die im Verfahren verwendete Approximation mit dem nichtäquidistanten Knotenvektor aus (3.4) erzielt dabei oft geringere Fehlerquadratsummen als die Approximation mit dem äquidistanten Knotenvektor. Insbesondere wenn die Vorgabepunkte große „Krümmungsänderungen“ wie im Fall der M-förmig vorgegebenen Punkte aufweisen, zeigt die nichtäquidistante Approximation ihre Vorteile: Zum Beispiel die rationale Approximation mit Feh-

Segmentanz.	$s_0 = 11$	$s_1 = 10$	$s_2 = 9$	$s_3 = 8$	$s_4 = 7$	$s_5 = 6$	$s_6 = 5$	$s_7 = 6$
F_1 (mm ²)	0.71	4.17	3.64	28.3	0.09	22.7	45.1	22.7
F_2 (mm ²)	0.047	4.21	7.42	3.27	0.07	3.61	38.7	3.61

Tabelle 3.11: Iteration der Segmentanzahl mit rationalen B-Splines bei M-förmigen Punkten und $\varepsilon = 5.0$ mm²

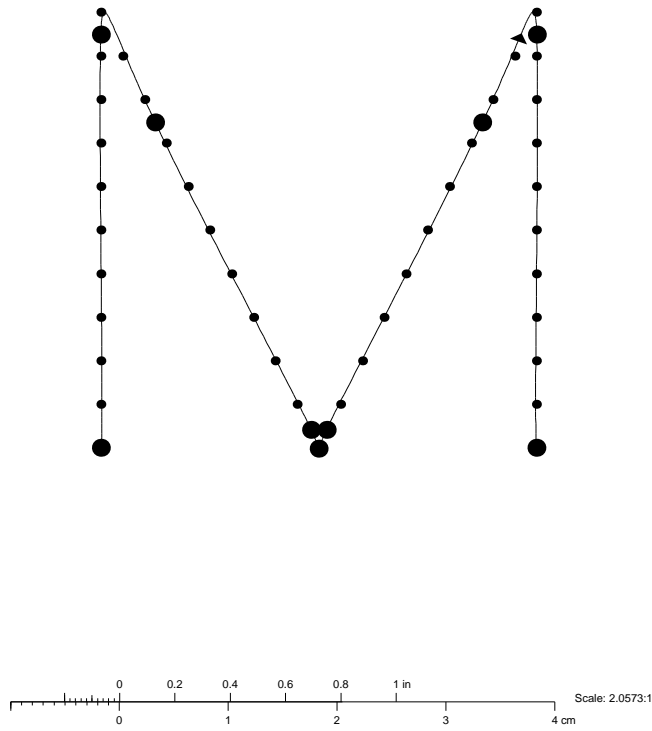


$\mathbf{T}' = (0, \dots, 0, 0.216, 0.252, 0.296, 0.424, 0.498, 0.499, 0.576, 0.704, 0.748, 0.784, 1, \dots, 1)$

Abbildung 3.6: Approximation von M-förmig gelegenen Punkten. Nichtäquidistante rat. B-Splinekurve mit 11 Segmenten, Fehlerquadratsumme 0.047 mm² und max. Fehler 0.12 mm (fett gezeichnete Punkte sind die den Knoten entsprechenden Kurvenpunkte)

Segmentanz.	$s_0 = 11$	$s_1 = 10$	$s_2 = 11$
F_1 (mm ²)	0.16	18.9	0.16
F_2 (mm ²)	2.49	6.42	2.49

Tabelle 3.12: Iteration der Segmentanzahl mit integralen B-Splines bei M-förmigen Punkten und $\varepsilon = 5.0$ mm²



$$\mathbf{T}' = (0, \dots, 0, 0.206, 0.334, 0.460, 0.500, 0.540, 0.666, 0.794, 1, \dots, 1)$$

Abbildung 3.7: Approximation von M-förmig gelegenen Punkten. Nichtäquidistante int. B-Splinekurve mit 8 Segmenten, Fehlerquadratsumme 0.42 mm^2 und max. Fehler 0.25 mm (fett gezeichnete Punkte sind die den Knoten entsprechenden Kurvenpunkte)

lertoleranz $\varepsilon = 5.0 \text{ mm}^2$. Hier hat die Kurve mit 6 Segmenten und mit äquidistantem Knotenvektor eine Fehlerquadratsumme von 22.7 mm^2 und einen maximalen Lotfehler von 2.88 mm , wohingegen die Kurve mit gleicher Segmentanzahl und mit nichtäquidistantem Knotenvektor eine Fehlerquadratsumme von 3.61 mm^2 und einen maximalen Lotfehler von 0.73 mm hat. Ein weiterer positiver Aspekt der nichtäquidistanten Approximation ist, daß der zusätzliche Rechenaufwand in (3.4) zur Bestimmung des Knotenvektors im Verhältnis zur Bestimmung der Approximationskurve vernachlässigbar ist.

Insgesamt erweist sich das Iterationsverfahren als eine effektive Methode, um eine kleine Segmentanzahl bei der Approximation mit B-Splinekurven (insbesondere rationaler Kurven) zu finden.

Auch in einem speziellen Fall der Approximation mit integralen oder rationalen B-Splineflächen (s. Abschnitt 2.1) ist eine Anwendung des Iterationsverfahrens zur Bestimmung der Segmentanzahl im Unterschied zu den Kurven aber nur mit äquidistanten Knotenvektoren \mathbf{U} , \mathbf{V} denkbar: Besitzen nämlich die Punkte \mathbf{P}_i eine Rechteckgitterstruktur, dann kann für jeden Gitterfaden in u - bzw. v -Richtung durch einen Iterationsschritt von (3.3) ein Wert für die Schrittweite Δs_u in u -Richtung bzw. Δs_v in v -Richtung bestimmt werden. Nach einem Kriterium, z. B. $\min\{\max\{\Delta s_u\}, \max\{\Delta s_v\}\}$, wird entweder ein Δs_u oder ein Δs_v ausgewählt und die entsprechende Segmentanzahl entweder in u - oder v -Richtung verändert. Anschließend wird wieder für jeden Gitterfaden eine Schrittweite Δs_u bzw. Δs_v mit (3.3) festgelegt usw. .

Kapitel 4

Approximation von Flächen mit Trimmkurven

Als Teilproblem bei der Approximation einer Fläche, d. h. abgetasteter Punkte der Fläche, durch eine B-Splinefläche tritt öfter der Fall auf, daß die gegebene Fläche Aussparungen (z. B. Bohrlöcher) enthält oder Randverläufe der Approximationsfläche durch Konstruktionsvorgaben geändert werden müssen. Um diesen Fall zu behandeln, werden zuerst Punkte der gegebenen Fläche mit der B-Splinefläche approximiert und danach die Bereiche der Approximationsfläche, die nicht benötigt werden oder außerhalb des neuen Randverlaufs der Approximationsfläche liegen, weggeschnitten. Die Kurven, die die Schnittländer beschreiben, bezeichnet man als *Trimmkurven*. Mathematisch läßt sich das Problem der Approximation mit Trimmkurven so beschreiben: Gegeben sind $m + 1$ Flächenpunkte $\mathbf{P}_i = \mathbf{Y}(u_i, v_i) \in \mathbb{R}^3$ auf der rationalen B-Splinefläche der Ordnungen p, q

$$\mathbf{Y}(u, v) := \frac{\sum_{j=0}^n \sum_{l=0}^r h_{jl} \mathbf{d}_{jl} N_{jp}(u) N_{lq}(v)}{\sum_{j=0}^n \sum_{l=0}^r h_{jl} N_{jp}(u) N_{lq}(v)} \quad , \quad \mathbf{d}_{jl} \in \mathbb{R}^3$$

Gesucht ist die Trimmkurve (Flächenkurve) $\mathbf{Y}(\mathbf{C}(t))$ mit der rationalen B-Splinekurve der Ordnung k

$$\mathbf{C}(t) := \begin{pmatrix} u(t) \\ v(t) \end{pmatrix} = \frac{\sum_{j=0}^n h_j \mathbf{d}_j N_{jk}(t)}{\sum_{j=0}^n h_j N_{jk}(t)} \quad , \quad \mathbf{d}_j := \begin{pmatrix} d_{jx} \\ d_{jy} \end{pmatrix} \in \mathbb{R}^2,$$

die die Punkte \mathbf{P}_i approximiert.

Die bisher angewendete Methode zur Approximation mit Trimmkurven [12], [13] basiert auf der Minimierung von

$$\sum_{i=0}^m d(t'_i) := \sum_{i=0}^m (u(t'_i) - u_i)^2 + (v(t'_i) - v_i)^2 \quad (4.1)$$

in der u - v -Parameterebene. Ausgehend von einer zentripetalen Parametrisierung der Startparameterwerte t_i bezüglich der $(u_i, v_i)^\top$ werden während des Approximationsprozesses iterativ die Parameterwerte t'_i gemäß

$$d(t'_i) \longrightarrow \min \quad (i = 0, \dots, m)$$

oder

$$(\mathbf{Y}(\mathbf{C}(t'_i)) - \mathbf{P}_i)^2 \longrightarrow \min$$

korrigiert. Die Zielfunktion (4.1) hat den Nachteil, daß in sie die Parameterdarstellung $\mathbf{Y}(u, v)$ der Fläche nicht eingeht. Da man aber an der Minimierung von $\mathbf{Y}(\mathbf{C}(t_i)) - \mathbf{P}_i$ interessiert ist, sollte $\mathbf{Y}(u, v)$ in geeigneter Weise berücksichtigt werden. Das nun zu präsentierende Verfahren tut dies und wendet bei der Approximation die neue Methode aus Abschnitt 1.3 an.

4.1 Approximation mit neuer Zielfunktion

Ziel der Approximation mit Trimmkurven soll die Minimierung des Fehlers zwischen $\mathbf{Y}(\mathbf{C}(t_i))$ und $\mathbf{P}_i = \mathbf{Y}(u_i, v_i)$ sein. Wir setzen voraus, daß $\mathbf{Y} \in C^2$ und $\partial \mathbf{Y} / \partial u \times \partial \mathbf{Y} / \partial v \neq \mathbf{0}$ ist. Deshalb betrachten wir

$$(\mathbf{Y}(\mathbf{C}(t_i)) - \mathbf{P}_i)^2 \quad (4.2)$$

Die Taylorentwicklung von $\mathbf{Y}(u, v)$ im Punkt $(u_i, v_i)^\top$ bis zu Gliedern 1. Ordnung ergibt

$$\mathbf{Y}(u, v) = \mathbf{Y}(u_i, v_i) + \mathbf{Y}_u(u_i, v_i) (u - u_i) + \mathbf{Y}_v(u_i, v_i) (v - v_i) + \mathbf{R}_2(u, v) \quad (4.3)$$

$\mathbf{R}_2(u, v)$ ist das quadratische Restglied mit

$$\lim_{(u, v) \rightarrow (u_i, v_i)} \mathbf{R}_2(u, v) = \mathbf{0}$$

und $\mathbf{Y}_u, \mathbf{Y}_v$ sind die partiellen Ableitungen von \mathbf{Y} nach u, v . Wird (4.3) in (4.2) eingesetzt, wobei das Restglied $\mathbf{R}_2(u, v)$ vernachlässigt wird, bekommt man

$$\begin{aligned} Fd &:= (\mathbf{Y}_u(u_i, v_i) (u(t_i) - u_i) + \mathbf{Y}_v(u_i, v_i) (v(t_i) - v_i))^2 = \\ &\mathbf{Y}_u^2(u_i, v_i) (u(t_i) - u_i)^2 + 2\mathbf{Y}_u(u_i, v_i)\mathbf{Y}_v(u_i, v_i) (u(t_i) - u_i)(v(t_i) - v_i) \\ &\quad + \mathbf{Y}_v^2(u_i, v_i) (v(t_i) - v_i)^2 \end{aligned} \quad (4.4)$$

Das Weglassen von $\mathbf{R}_2(u, v)$ laßt sich dadurch rechtfertigen, daß ein kleiner Fehler (4.2) (Ziel einer guten Approximation) auch einen kleinen Wert von $\mathbf{C}(t_i) - (u_i, v_i)^\top$ zur Folge hat. (4.4) stellt aus differentialgeometrischer Sicht gerade die *erste Fundamentalform* dar, wenn man die Differenzen $u(t_i) - u_i, v(t_i) - v_i$ durch Differentiale ersetzt. Wenn die u - und v -Parameterlinien von \mathbf{Y} orthogonal zueinander sind, gilt insbesondere

$$Fd = \mathbf{Y}_u^2(u_i, v_i) (u(t_i) - u_i)^2 + \mathbf{Y}_v^2(u_i, v_i) (v(t_i) - v_i)^2$$

Im Fall, daß $\mathbf{C}(t)$ eine integrale B-Splinekurve ist ($h_j = 1$), führt die Minimierung von $\sum_{i=0}^m Fd$ auf lineares Gleichungssystem für die Unbekannten \mathbf{d}_j . Um auch bei der Minimierung mit einer rationalen B-Splinekurve $\mathbf{C}(t)$ ein lineares Gleichungssystem für die

Unbekannten \mathbf{d}_j , h_j zu erhalten, genügt (4.4) noch nicht. Deshalb wird (4.4) nach oben abgeschätzt durch

$$Fd \leq 2 (\mathbf{Y}_u^2(u_i, v_i) (u(t_i) - u_i)^2 + \mathbf{Y}_v^2(u_i, v_i) (v(t_i) - v_i)^2) \quad (4.5)$$

Die rechte Seite von (4.5) läßt sich als Fehlerquadrat der Punkte

$$\mathbf{x}(t_i) := \sqrt{2} \begin{pmatrix} |\mathbf{Y}_u(u_i, v_i)| u(t_i) \\ |\mathbf{Y}_v(u_i, v_i)| v(t_i) \end{pmatrix}$$

und

$$\mathbf{p}_i := \sqrt{2} \begin{pmatrix} |\mathbf{Y}_u(u_i, v_i)| u_i \\ |\mathbf{Y}_v(u_i, v_i)| v_i \end{pmatrix}$$

deuten. Nun können wir die Ergebnisse aus Kapitel 1 auf die Approximation mit Trimmkurven übertragen: In Abschnitt 1.2 hatten wir die Zielfunktion

$$\bar{F} = \sum_{i=0}^m \bar{\mathbf{X}}^2(t_i) - (\bar{\mathbf{X}}(t_i) \cdot \bar{\mathbf{P}}_i)^2 + \lambda \sum_{i=0}^m \left(\sum_{j=0}^n h_j N_{jk}(t_i) - 1 \right)^2$$

zur Approximation von $m+1$ Punkten $\mathbf{P}_i \in \mathbb{R}^3$ und Parameterwerten t_i mit einer rationalen B-Splinekurve $\mathbf{X}(t)$ verwendet, wobei

$$\bar{\mathbf{X}}(t) := \begin{pmatrix} \sum_{j=0}^n h_j N_{jk}(t) \\ \sum_{j=0}^n \bar{d}_{jx} N_{jk}(t) \\ \sum_{j=0}^n \bar{d}_{jy} N_{jk}(t) \\ \sum_{j=0}^n \bar{d}_{jz} N_{jk}(t) \end{pmatrix}, \quad \bar{\mathbf{P}}_i := \frac{1}{\sqrt{1 + p_{ix}^2 + p_{iy}^2 + p_{iz}^2}} \begin{pmatrix} 1 \\ p_{ix} \\ p_{iy} \\ p_{iz} \end{pmatrix}$$

homogene Koordinaten von $\mathbf{X}(t)$ sowie \mathbf{P}_i sind. Analog wird jetzt die Zielfunktion

$$\bar{F} = \sum_{i=0}^m \bar{\mathbf{x}}^2(t_i) - (\bar{\mathbf{x}}(t_i) \cdot \bar{\mathbf{p}}_i)^2 + \lambda \sum_{i=0}^m \left(\sum_{j=0}^n h_j N_{jk}(t_i) - 1 \right)^2 \quad (4.6)$$

mit

$$\bar{\mathbf{x}}(t_i) := \begin{pmatrix} \sum_{j=0}^n h_j N_{jk}(t_i) \\ |\mathbf{Y}_u(u_i, v_i)| \sum_{j=0}^n \bar{d}_{jx} N_{jk}(t_i) \\ |\mathbf{Y}_v(u_i, v_i)| \sum_{j=0}^n \bar{d}_{jy} N_{jk}(t_i) \end{pmatrix}, \quad \bar{\mathbf{p}}_i := c_i \begin{pmatrix} 1 \\ |\mathbf{Y}_u(u_i, v_i)| u_i \\ |\mathbf{Y}_v(u_i, v_i)| v_i \end{pmatrix}$$

und

$$c_i := \frac{1}{\sqrt{1 + (|\mathbf{Y}_u(u_i, v_i)| u_i)^2 + (|\mathbf{Y}_v(u_i, v_i)| v_i)^2}}$$

benutzt. Bei $\bar{\mathbf{x}}(t_i)$ und $\bar{\mathbf{p}}_i$ sind die Faktoren $\sqrt{2}$ weggelassen, die bei $\mathbf{x}(t_i)$ und \mathbf{p}_i auftreten, da sie keinen Einfluß auf das Minimierungsergebnis von (4.6) haben. Unbekannte der Optimierung sind die h_j , $\bar{d}_{jx} := h_j d_{jx}$, $\bar{d}_{jy} := h_j d_{jy}$.

Die notwendigen Bedingungen zur Minimierung von (4.6) ergeben ein lineares Gleichungssystem ähnlich zu (1.18). Die Aussagen und Sätze aus Abschnitt 1.3 gelten für das Lösen des Gleichungssystems analog. Zur Verringerung des Zielfunktionswertes werden ebenfalls die Parameterwerte t_i korrigiert (vgl. (1.30)). Die Startparameterwerte t_i können als

zentripetale oder chordale Parametrisierung der Raumpunkte \mathbf{P}_i , s. [11], oder in Anlehnung an diese Parametrisierungen zum Beispiel als zentripetale Parametrisierung von Flächenpunkten $\mathbf{P}_i = \mathbf{Y}(u_i, v_i)$ so bestimmt werden

$$t_i = \frac{\sum_{l=1}^i \sqrt[4]{\mathbf{Y}_u^2 (u_l - u_{l-1})^2 + 2\mathbf{Y}_u \mathbf{Y}_v (u_l - u_{l-1})(v_l - v_{l-1}) + \mathbf{Y}_v^2 (v_l - v_{l-1})^2}}{s}$$

$$s = \sum_{l=1}^m \sqrt[4]{\mathbf{Y}_u^2 (u_l - u_{l-1})^2 + 2\mathbf{Y}_u \mathbf{Y}_v (u_l - u_{l-1})(v_l - v_{l-1}) + \mathbf{Y}_v^2 (v_l - v_{l-1})^2}.$$

4.2 Testresultate

4.2.1 Approximationsverhalten des Verfahrens

Das Approximationsverhalten des neuen Verfahrens wird durch einige Beispiele illustriert. In den Beispielen wurde mit Trimmkurven auf bikubischen rationalen B-Splineflächen approximiert. Die Kurven $\mathbf{C}(t)$ in der u - v -Parameterebene der Fläche sind offene kubische rationale B-Splinekurven mit äquidistantem Knotenvektor. Jedes Beispiel besteht aus zwei Abbildungen. Die eine Abbildung zeigt die Trimmkurve $\mathbf{Y}(\mathbf{C}(t))$ und die Vorgabepunkte $\mathbf{Y}(u_i, v_i)$ auf der gegebenen Fläche, die andere die Kurve $\mathbf{C}(t)$ und die Punkte $(u_i, v_i)^\top$ in der u - v -Parameterebene der Fläche. Die folgenden Fehlerwerte sind maximale Lotfehler zwischen der Trimmkurve und den vorgegebenen Flächenpunkten. Zur Einordnung der Größenverhältnisse zwischen der Trimmkurve und dem Fehlerwert ist in den Abbildungen ein Maßstab eingezeichnet.

Die ersten beiden Beispiele dienen dazu, den Einfluß der Parameterdarstellung $\mathbf{Y}(u, v)$ auch in der Parameterebene auf das Approximationsergebnis zu demonstrieren: Abbildung 4.1 zu Beispiel 1 stellt die Approximation von 41 Punkten auf einem Zylinder dar. In Abbildung 4.2 sieht man das dazugehörige Ergebnis in der u - v -Parameterebene. In der letzteren Abbildung liegen die Punkte auf einem Kreis. Die Kurve $\mathbf{C}(t)$ hat 1 Segment und die Trimmkurve approximiert die Punkte mit einem Fehler von 1.39 mm.

Im nächsten Beispiel (Abbildung 4.3 und 4.4) wurden 41 Punkte mit einem Fehler von 1.02 mm auf einer würfelförmigen Fläche approximiert. Die Kurve $\mathbf{C}(t)$ hat wie im ersten Beispiel 1 Segment und die Vorgabepunkte liegen in der u - v -Parameterebene ebenfalls auf einem Kreis mit gleichem Radius (Abbildung 4.4). Vergleicht man Abbildung 4.2 und 4.4 miteinander, so erkennt man deutlich, daß die Parameterdarstellung $\mathbf{Y}(u, v)$ der Fläche das Approximationsergebnis auch in der Parameterebene beeinflusst. Dies erklärt sich aus der Tatsache, daß in die Zielfunktion (4.6) die Parameterdarstellung der Fläche eingeht, was bei der Zielfunktion (4.1) jedoch nicht der Fall ist. Bei den beiden eben erläuterten Beispielen kann man durch Erhöhung der Segmentanzahl erheblich geringere Fehlerwerte erreichen, wodurch sich dann aber die Approximationskurven in der Parameterebene optisch kaum unterscheiden.

Als drittes Beispiel sind in Abbildung 4.5 41 Punkte auf einem Zylinder approximiert. In der dazugehörigen u - v -Parameterebene (Abbildung 4.6) sind diese Punkte sinusförmig gelegen. Die Kurve $\mathbf{C}(t)$ besitzt 5 Segmente und der Approximationsfehler beträgt 0.039 mm.

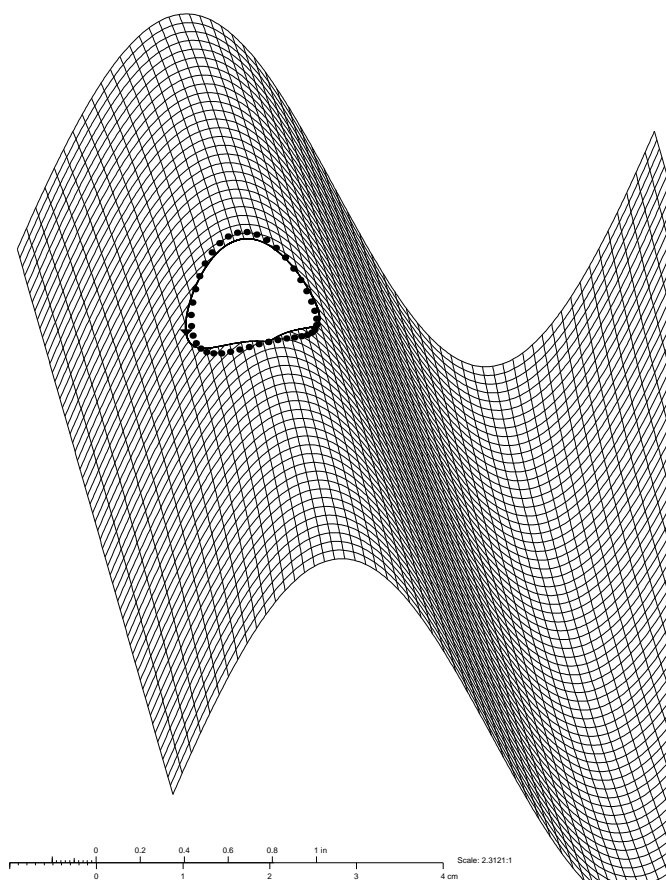


Abbildung 4.1: Trimmkurvenapproximation von Punkten auf einem Zylinder. Max. Fehler 1.39 mm

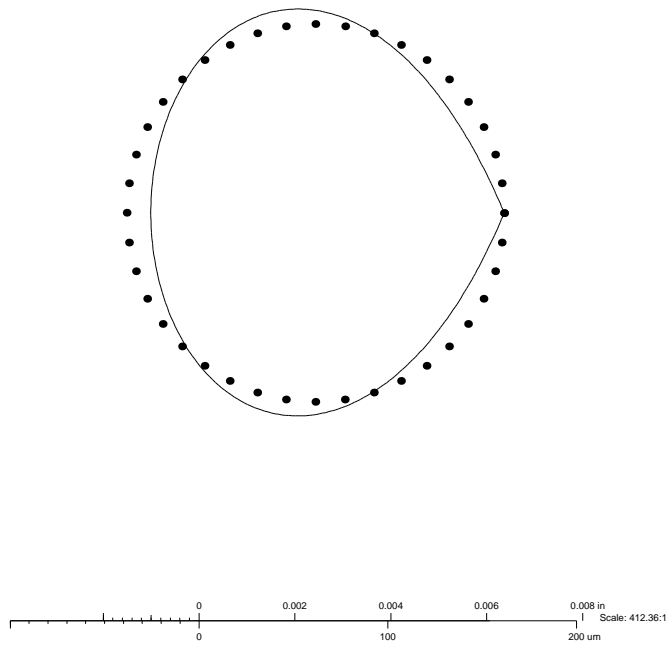


Abbildung 4.2: Trimmkurvenapproximation von Punkten des Zylinders in der u - v -Parameterebene auf einem Kreis

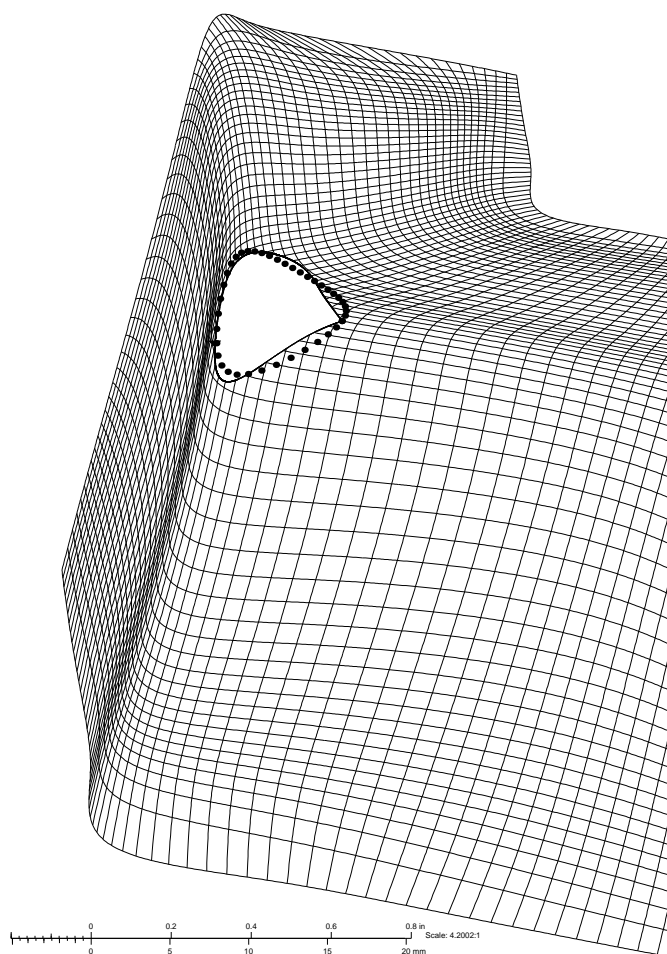


Abbildung 4.3: Trimmkurvenapproximation von Punkten auf einer würfelähnlichen Fläche. Max. Fehler 1.02 mm

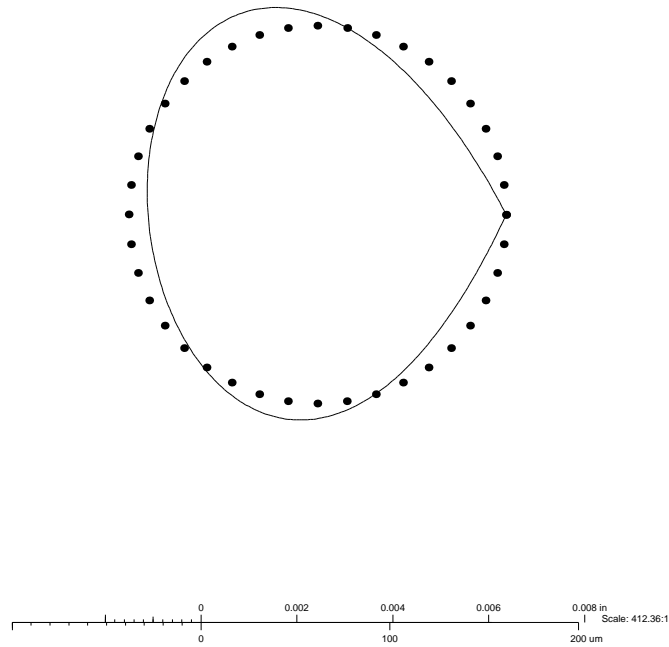


Abbildung 4.4: Trimmkurvenapproximation von Punkten der würfelähnlichen Fläche in der $u-v$ -Parameterebene auf einem Kreis

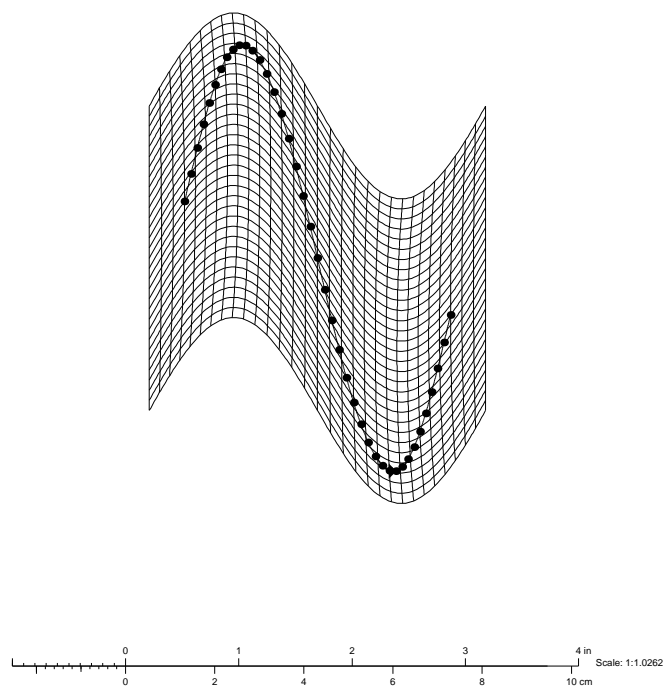


Abbildung 4.5: Trimmkurvenapproximation von Punkten auf einem Zylinder. Max. Fehler 0.039 mm

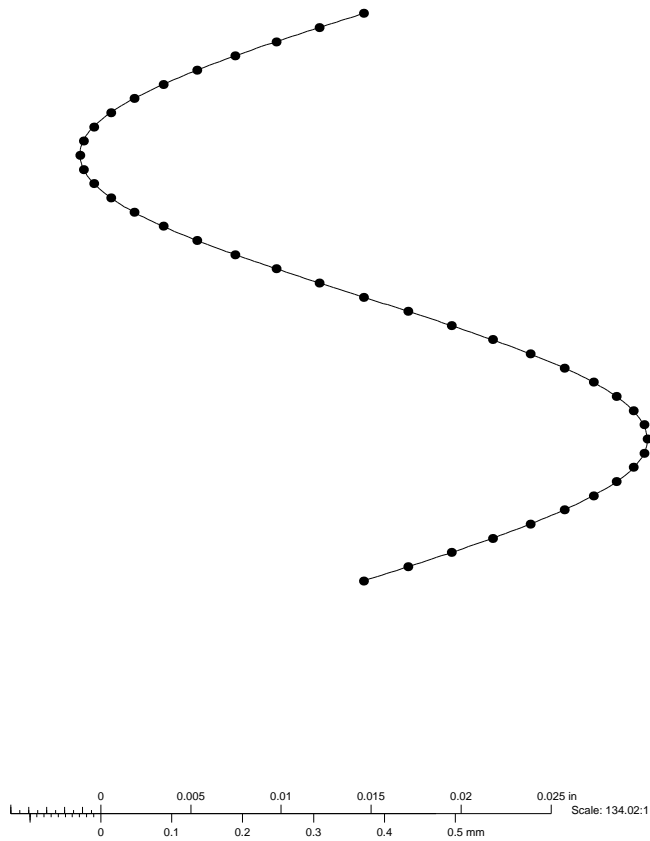


Abbildung 4.6: Trimmkurvenapproximation von sinusförmig gelegenen Punkten des Zylinders in der u - v -Parameterebene

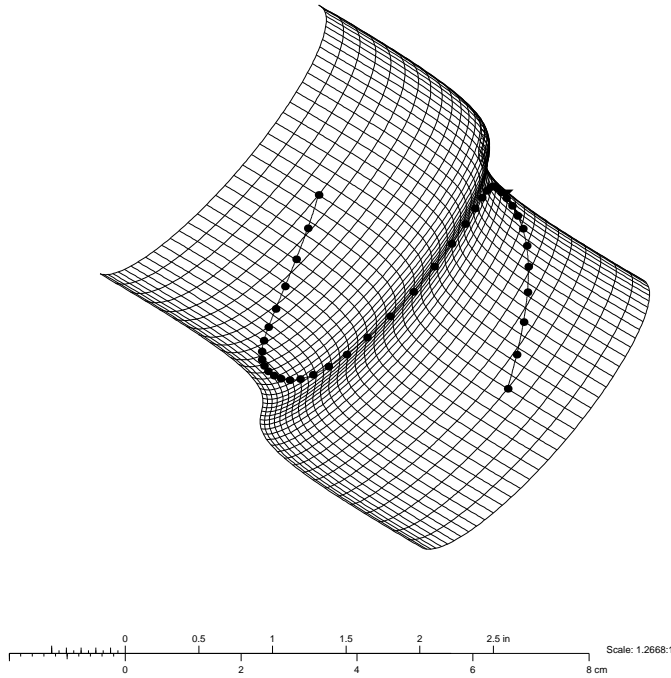


Abbildung 4.7: Trimmkurvenapproximation von Punkten auf einer Translationsfläche. Max. Fehler 0.064 mm

Schließlich zeigt Abbildung 4.7 die Approximation von 41 Punkten auf einer Translationsfläche. Wie in Abbildung 4.6 haben die Punkte in u - v -Parameterebene (Abbildung 4.8) die gleiche sinusförmige Anordnung. Die Segmentanzahl der Kurve $\mathbf{C}(t)$ ist ebenso wie im vorhergehenden Beispiel 5 und der Approximationsfehler der Trimmkurve ist 0.064 mm. Im Gegensatz zu den beiden ersten Beispielen sind im Vergleich der Abbildungen 4.6 und 4.8 optisch keine Unterschiede festzustellen.

Die Beispiele machen deutlich, daß die Verwendung der Zielfunktion (4.6) im Unterschied zu (4.1) auch das Approximationsergebnis in der u - v -Parameterebene beeinflusst, da in (4.6) über die partiellen Ableitungen $\mathbf{Y}_u, \mathbf{Y}_v$ die Parameterdarstellung $\mathbf{Y}(u, v)$ der Fläche eingeht, was in (4.1) nicht der Fall ist. Außerdem demonstrieren die beiden letzten Beispiele (Abbildungen 4.5, 4.6, 4.7 und 4.8), daß das neue Verfahren sehr gute Approximationsergebnisse ermöglicht.

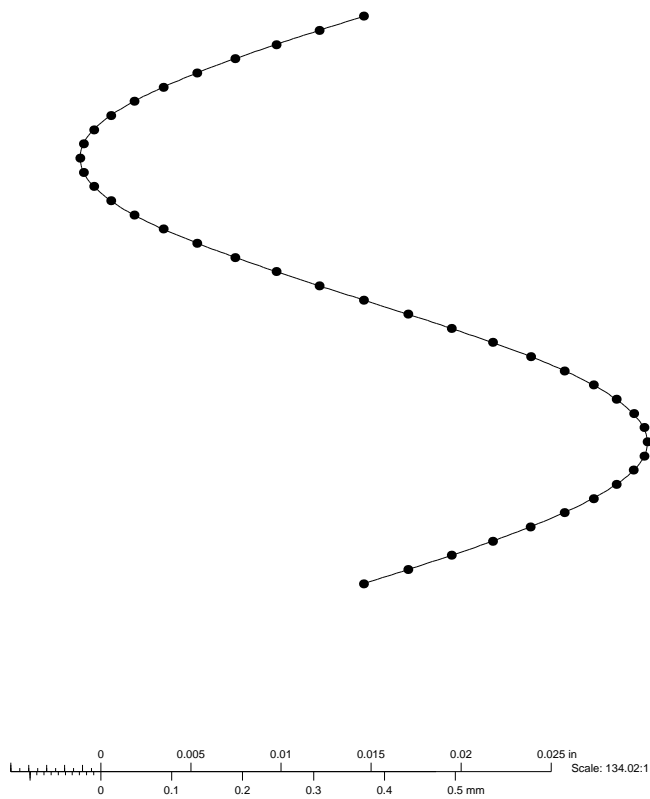


Abbildung 4.8: Trimmkurvenapproximation von sinusförmig gelegenen Punkten der Translationsfläche in der u - v -Parameterebene

4.2.2 Verfahren im Vergleich

Um die Qualität des neuen Verfahrens zu beurteilen, wird es mit der bisher verwendeten Methode (Minimierung von (4.1) und Parameterkorrektur durch $(\mathbf{Y}(\mathbf{C}(t'_i)) - \mathbf{P}_i)^2 \rightarrow \min$ ($i = 0, \dots, m$)) an einigen Beispielen verglichen. Während die Kurven $\mathbf{C}(t)$ in den u - v -Parameterebenen der Flächen beim neuen Verfahren in allen Beispielen offene rationale B-Splinekurven sind, sind sie in der Vergleichsmethode offene integrale B-Splinekurven ($h_j = 1$). Sowohl für die rationalen als auch für die integralen Kurven wurden immer äquidistante Knoten verwendet. Jedes Beispiel besteht wieder aus zwei Abbildungen, die die Approximationskurven des neuen Verfahrens zusammen mit den Vorgabepunkten zum einen auf der gegebenen Fläche, zum anderen in der Parameterebene der Fläche zeigen. Die Vorgabepunkte der Beispiele liegen auf den Flächen ungefähr gleichabständig. Die im folgenden angegebenen Fehlerwerte bezeichnen maximale Lotfehler zwischen der Trimmkurve $\mathbf{Y}(\mathbf{C}(t))$ und den Vorgabepunkten $\mathbf{Y}(u_i, v_i)$. Um die Fehlerwerte in Bezug auf die Ausdehnung der Trimmkurve einschätzen zu können, enthalten die Abbildungen Maßstäbe.

Das erste Beispiel stellt die Approximation von 97 Punkten auf einer biquadratischen integralen B-Splinefläche (Abbildung 4.9) dar. In der u - v -Parameterebene liegen die Punkte sehr ungleichmäßig auf einem Kreis (Abbildung 4.10). Der Fehler der kubischen rationalen Trimmkurve mit 16 Segmenten beträgt 0.098 mm, der Fehler der kubischen integralen Trimmkurve mit 32 Segmenten 0.1 mm. Die Berechnung mit dem neuen Verfahren benötigte bei 20 Iterationen (Parameterkorrekturschritte) 9 s, mit dem Vergleichsverfahren bei 14 Iterationen 11 s. Besonders ins Auge fällt, daß das neue Verfahren nur die Hälfte der Segmente der Vergleichsmethode braucht, um einen ähnlichen Fehler zu erzielen.

In Abbildung 4.11 zu Beispiel 2 sind 88 Punkte auf einem Parabelbogen wieder sehr ungleichmäßig in der u - v -Parameterebene vorgegeben, die mit einer quadratischen rationalen B-Splinekurve mit 16 Segmenten approximiert werden. Die entsprechende Approximationskurve mit den Vorgabepunkten auf der Fläche, die identisch mit der Fläche des vorhergehenden Beispiels ist, zeigt Abbildung 4.12. Die Rechenzeit des neuen Verfahrens belief sich bei 63 Iterationen auf 23 s, wobei sich ein Fehler von 0.091 mm ergab. In der Vergleichsmethode wurde mit einer quadratischen integralen Kurve mit 32 Segmenten und einer Rechenzeit von 5 s (7 Iterationen) ein Fehler von 0.093 mm erreicht. Auch hier ist die deutlich geringere Anzahl von Segmenten (Faktor 2!) des neuen Verfahrens bei etwa gleichen Fehlern hervorzuheben.

Als drittes Beispiel zeigt Abbildung 4.13 die Approximation von 48 rechteckförmig gelegenen Punkten in der Parameterebene und Abbildung 4.14 die Entsprechung auf einer bikubischen integralen B-Splinefläche. Der Fehler der kubischen rationalen Trimmkurve mit 11 Segmenten ist 0.049 mm. Die dazu notwendigen 30 Iterationen wurden in 6 s berechnet. Der Fehler der kubischen integralen Trimmkurve mit ebenfalls 11 Segmenten ist 0.05 mm, wobei 76 Iterationen mit 20 s Rechenzeit erforderlich waren.

Zum Schluß (Beispiel 4) werden in Abbildung 4.15 36 rautenförmig gelegene Punkte mit einer kubischen rationalen B-Splinekurve mit 10 Segmenten in der Parameterebene und in Abbildung 4.16 die dazugehörigen Punkte auf einer biquadratischen integralen B-Splinefläche approximiert. Die Berechnung mit dem neuen Verfahren benötigte 6 s für 44 Iterationen und der Fehler beträgt 0.049 mm. In der Vergleichsmethode ergab sich mit einer kubischen integralen B-Splinekurve mit 14 Segmenten ein Fehler von 0.05 mm. Die

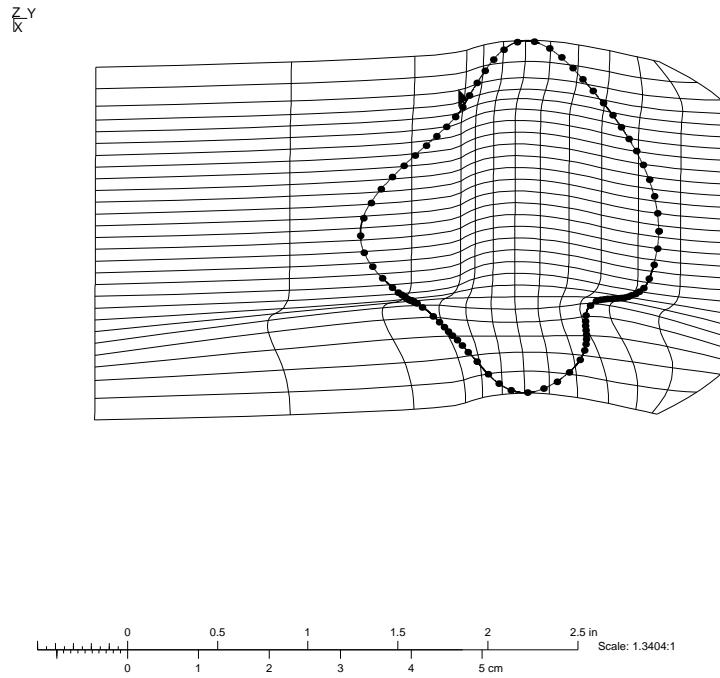


Abbildung 4.9: Trimmkurvenapproximation von Punkten auf der Fläche zu Beispiel 1. Max. Fehler 0.098 mm

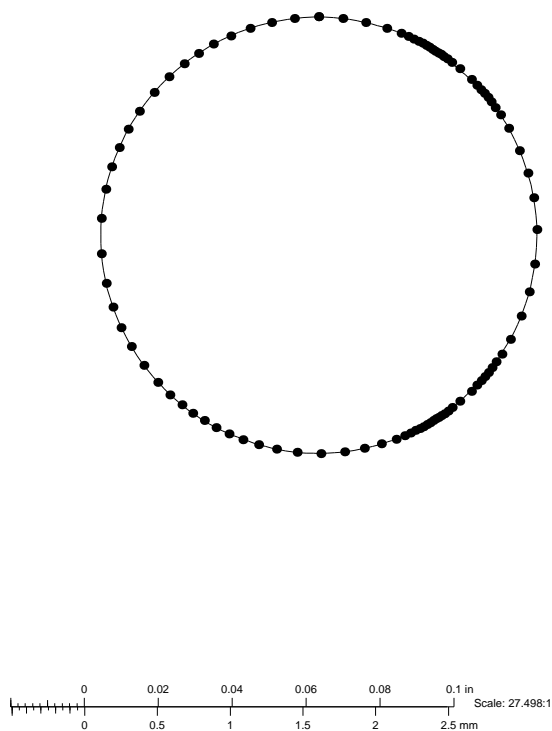


Abbildung 4.10: Trimmkurvenapproximation von Punkten auf einem Kreis in der $u-v$ -Parameterebene der Fläche zu Beispiel 1

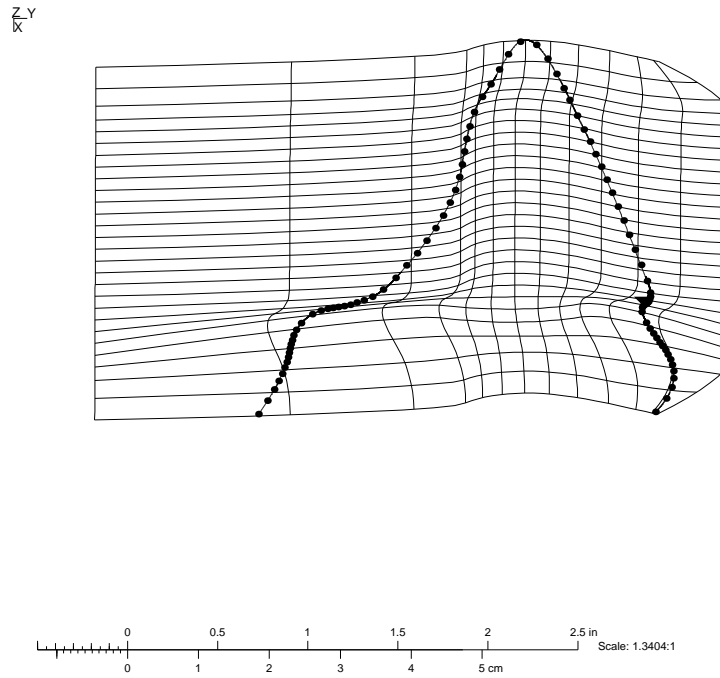


Abbildung 4.11: Trimmkurvenapproximation von Punkten auf der Fläche zu Beispiel 2. Max. Fehler 0.091 mm

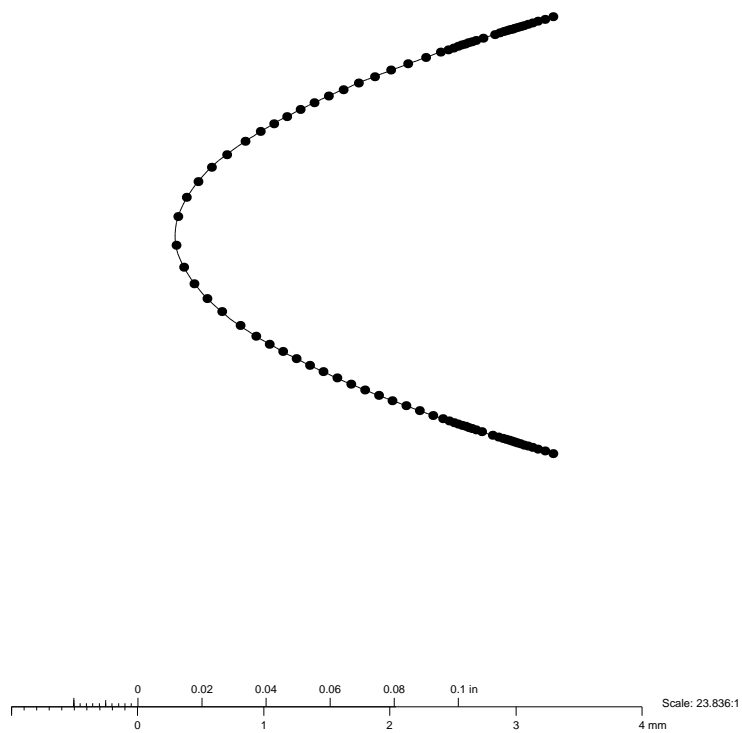


Abbildung 4.12: Trimmkurvenapproximation von Punkten auf einer Parabel in der u - v -Parameterebene der Fläche zu Beispiel 2

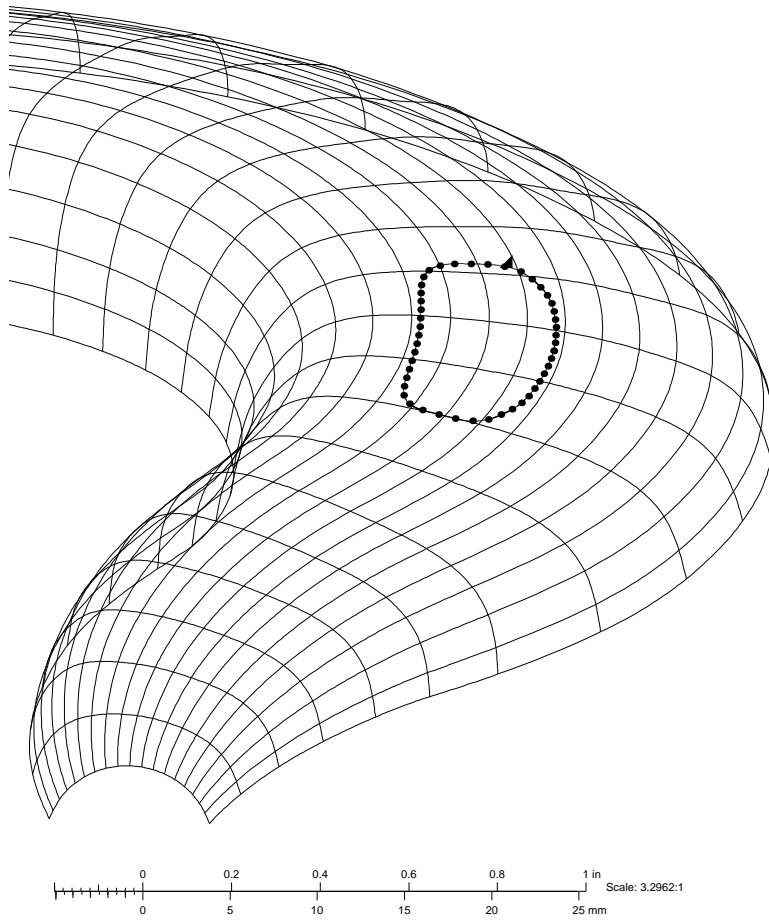


Abbildung 4.13: Trimmkurvenapproximation von Punkten auf der Fläche zu Beispiel 3. Max. Fehler 0.049 mm

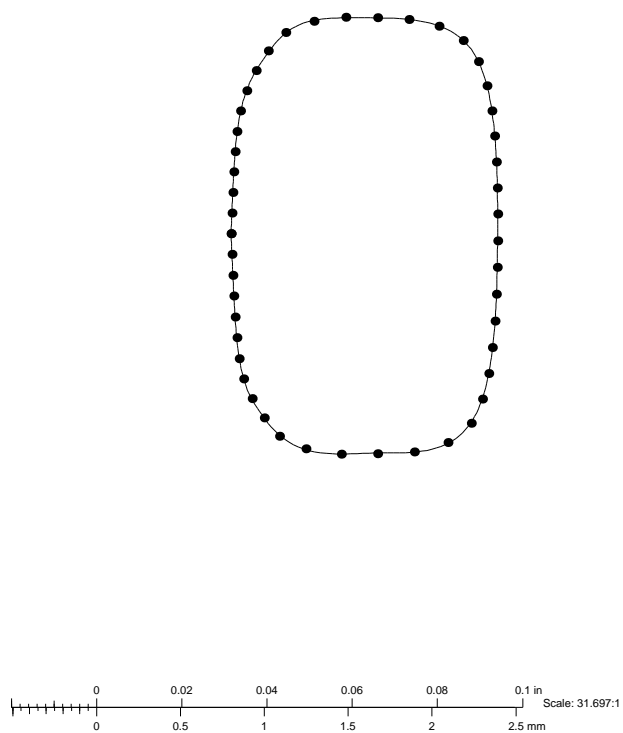


Abbildung 4.14: Trimmkurvenapproximation von rechteckförmig gelegenen Punkten in der u - v -Parameterebene der Fläche zu Beispiel 3

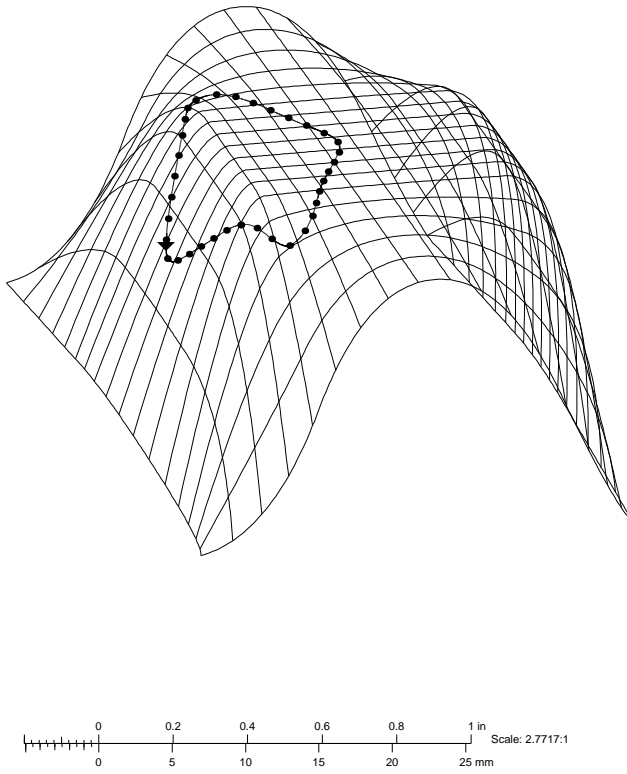


Abbildung 4.15: Trimmkurvenapproximation von Punkten auf der Fläche zu Beispiel 4. Max. Fehler 0.049 mm

Berechnung dauerte für 8 Iterationen 2 s. Hier fällt die Anzahl der Segmente des neuen Verfahrens bei etwa gleichen Fehlern wieder geringer aus als bei der Vergleichsmethode.

Die Beispiele zeigen, daß das neue Verfahren zum Erreichen ähnlicher Fehlerwerte teilweise deutlich weniger Segmente braucht als die Vergleichsmethode. Dies fällt besonders bei dem ersten und zweiten Beispiel auf. Der Unterschied bei der Anzahl der Segmente zwischen dem neuen Verfahren und der Vergleichsmethode ist teils auf die Verwendung ra-

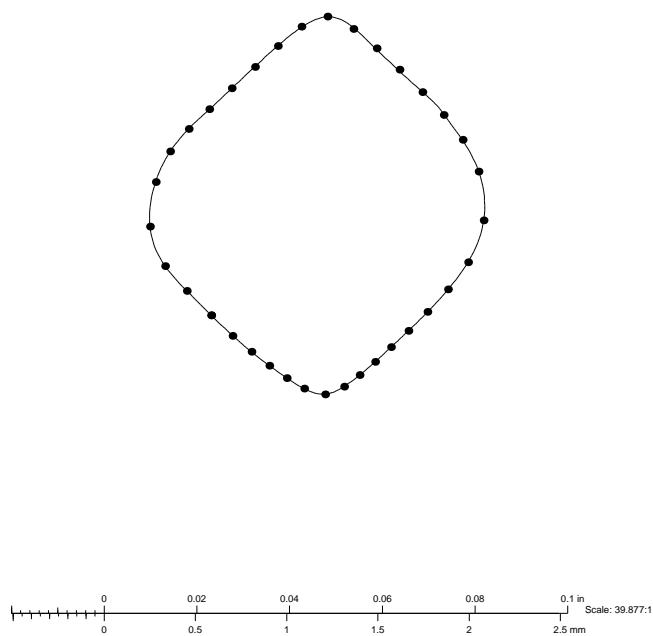


Abbildung 4.16: Trimmkurvenapproximation von rautenförmig gelegenen Punkten in der u - v -Parameterebene der Fläche zu Beispiel 4

tionaler B-Splinekurven gegenüber integralen B-Splinekurven, teils auf das neue Verfahren zurückzuführen. Letztere Aussage belegen beispielsweise die Approximationsergebnisse der Punkte dieser beiden Beispiele, die nochmal mit integralen B-Splinekurven unter Verwendung des neuen Verfahrens approximiert wurden, d. h. der Minimierung von (4.6) mit $h_j = 1$: In beiden Fällen hatten die integralen Kurven 26 Segmente mit Fehlern von 0.095 mm bzw. 0.100 mm und somit 6 Segmente weniger als die entsprechenden integralen Kurven, die sich bei der Minimierung von (4.1) ergaben (s. Beispiel 1 und 2). Diese Ergebnisse sind durch folgende Überlegung gerechtfertigt: In den Beispielen sind die Vorgabepunkte $\mathbf{Y}(u_i, v_i)$ auf der Fläche ungefähr gleichabständig und dicht gelegen, d. h.

$$(\mathbf{Y}(u_i, v_i) - \mathbf{Y}(u_{i-1}, v_{i-1}))^2 \approx c, \quad c = \text{const.} \quad (4.7)$$

Taylorentwicklung von $\mathbf{Y}(u, v)$ in $(u_i, v_i)^\top$ bis zu Gliedern 1. Ordnung ergibt (vgl. (4.4)) bei (4.7)

$$\begin{aligned} (\mathbf{Y}() - \mathbf{Y}())^2 &\approx \mathbf{Y}_u^2() (u_{i-1} - u_i)^2 + 2\mathbf{Y}_u()\mathbf{Y}_v() (u_{i-1} - u_i)(v_{i-1} - v_i) \\ &\quad + \mathbf{Y}_v^2() (v_{i-1} - v_i)^2 \approx c \end{aligned} \quad (4.8)$$

(4.8) bedeutet, daß bei kleinerem Abstand zwischen Punkten $(u_i, v_i)^\top$ und $(u_{i-1}, v_{i-1})^\top$ in der Parameterebene die Beträge der partiellen Ableitungen $|\mathbf{Y}_u(u_i, v_i)|$, $|\mathbf{Y}_v(u_i, v_i)|$ größere Werte annehmen und umgekehrt bei größerem Abstand zwischen den Punkten $|\mathbf{Y}_u(u_i, v_i)|$, $|\mathbf{Y}_v(u_i, v_i)|$ kleinere Werte haben.

In der Zielfunktion (4.6) sind die homogenen Darstellungen $\bar{\mathbf{x}}(t_i)$, $\bar{\mathbf{p}}_i$ mit den $|\mathbf{Y}_u(u_i, v_i)|$, $|\mathbf{Y}_v(u_i, v_i)|$ gewichtet, so daß in Bereichen der Parameterebene mit kleinerem Abstand der Punkte $(u_i, v_i)^\top$ und damit größeren Werten von $|\mathbf{Y}_u(u_i, v_i)|$, $|\mathbf{Y}_v(u_i, v_i)|$ ein kleinerer Approximationsfehler $|\mathbf{C}(t_i) - (u_i, v_i)^\top|$ zu erwarten ist als in Bereichen mit größerem Abstand der Punkte $(u_i, v_i)^\top$. Dies ist auch sinnvoll, denn die in den Abbildungen 4.9, 4.11 gezeichneten u - und v -Parameterlinien liegen in der Parameterebene gleichabständig, d. h. liegt ein Flächenpunkt $\mathbf{Y}(u_i, v_i)$ in einem Bereich der Fläche, in der die Parameterlinien größere Abstände haben, dann ist der Approximationsfehler $|\mathbf{Y}(\mathbf{C}(t_i)) - \mathbf{Y}(u_i, v_i)|$ auf der Fläche größer als der Approximationsfehler $|\mathbf{Y}(\mathbf{C}(t_k)) - \mathbf{Y}(u_k, v_k)|$ auf der Fläche für einen Flächenpunkt $\mathbf{Y}(u_k, v_k)$, der in einem Flächenbereich mit kleineren Abständen der Parameterlinien liegt, wobei die Approximationsfehler $|\mathbf{C}(t_i) - (u_i, v_i)^\top|$, $|\mathbf{C}(t_k) - (u_k, v_k)^\top|$ in der Parameterebene gleich sind. Nun liegen aber gerade die Punkte $(u_i, v_i)^\top$ in Bereichen der Parameterebene dichter, wo die entsprechenden Flächenbereiche Parameterlinien mit größeren Abständen haben. Dagegen berücksichtigt die Vergleichsmethode diesen Aspekt nicht, da in (4.1) nur die Punkte $(u_i, v_i)^\top$ der Parameterebene eingehen.

In Bezug auf die Rechenzeit erweist sich das neue Verfahren nicht wesentlich langsamer als die Vergleichsmethode. Insgesamt gesehen stellt das neue Verfahren daher eine wesentliche Verbesserung gegenüber der bisherigen Methode dar.

Schlußwort

In dieser Arbeit wurden neue Verfahren zur Approximation mit rationalen B-Splinekurven und -flächen, zur Bestimmung einer geeigneten Segmentanzahl für rationale B-Splinekurven, um einen vorgegebenen Approximationsfehler einzuhalten, und zur Approximation mit rationalen Trimmkurven vorgestellt.

Die Verfahren zur Approximation mit rationalen B-Splinekurven und -flächen sowie mit rationalen Trimmkurven stellen, wie die angeführten Beispiele zeigen, eine erhebliche Verbesserung gegenüber den bisher bekannten Methoden zur Lösung dieser Approximationsprobleme dar. Im einzelnen sind dies: die Lösung jeweils nur linearer Systeme (geringer Rechenaufwand), geringer Speicherverbrauch, keine Beschränkung der Ordnung der B-Splinekurven und -flächen, Gewährleistung positiver Gewichte und die Berücksichtigung der Parameterdarstellung der Fläche $\mathbf{Y}(u, v)$ durch Wichtung der Zielfunktion (4.6) mit den partiellen Ableitungen $|\mathbf{Y}_u()|$, $|\mathbf{Y}_v()|$.

Auch das Iterationsverfahren, das die Segmentanzahl von B-Splinekurven ermittelt, erweist sich durch die Verwendung des abgewandelten Sekantenverfahrens als wirkungsvoller Lösungsansatz zu dieser Problemstellung.

Alle präsentierten Verfahren eignen sich auch wegen der in den Testbeispielen gezeigten Stabilität zum künftigen Einsatz in einem CAD-System, das rationale B-Splinekurven und -flächen verwendet, da von weiteren Steigerungen der Rechnerleistungen ausgegangen werden kann.

Aus dieser Arbeit heraus ergeben sich weiterführende Fragestellungen, z. B.: Es wäre interessant zu untersuchen, ob sich analog zur integralen Flächenapproximation, bei der die Zielfunktion (2.4) minimiert wird, bei der rationalen Flächenapproximation auch ein Energieterm zur Glättung der Fläche zur Zielfunktion (2.2) hinzufügen läßt, so daß ein lineares Optimierungsproblem erhalten bleibt. Der Energieterm soll dabei wie im integralen Fall näherungsweise die Biegeenergie der dünnen Platte beschreiben und die de Boor-Punkte sowie die Gewichte als Optimierungsvariablen enthalten.

Literaturverzeichnis

- [1] Boehm, W. & Prautzsch, H.: Numerical Methods (1993), A. K. Peters.
- [2] Cox, M. G.: Practical spline approximation. Topics in Numerical Analysis, Lecture Notes in Mathematics 965 (ed. P. R. Turner) (1981), 79–112.
- [3] Dierckx, P.: Curve and surface fitting with splines (1993), Clarendon Press.
- [4] Dietz, U.: Fair Surface Reconstruction from Point Clouds, in Mathematical Methods for Curves and Surfaces II, M. Dæhlen, T. Lyche, and L. L. Schumaker (eds.) (1998), Vanderbilt University Press, 79–86.
- [5] Eck, M. & Hadenfeld, J.: Knot removal for parametric B-spline curves, Computer Aided Geometric Design 12 (1995), 259–282.
- [6] Elsässer, B. & Hoschek, J.: Approximation of digitized points by surfaces of revolution, Computer & Graphics Vol. 20 No. 1 (1996), 85–94.
- [7] Elsässer, B.: Approximation with Rational B-spline Curves and Surfaces, in Mathematical Methods for Curves and Surfaces II, M. Dæhlen, T. Lyche, and L. L. Schumaker (eds.) (1998), Vanderbilt University Press, 87–94.
- [8] Farin, G. E.: NURBS curves and surfaces: from projective geometry to practical use (1995), A. K. Peters.
- [9] Finck von Finckenstein, K. Graf: Einführung in die Numerische Mathematik (1977), Hanser.
- [10] Giering, O.: Vorlesungen über höhere Geometrie (1982), Vieweg.
- [11] Hoschek, J. & Lasser, D.: Fundamentals of Computer Aided Geometric Design (1993), A. K. Peters.
- [12] Hoschek, J. & Schneider, F.-J.: Approximate Spline Conversion for Integral and Rational Bézier and B-Spline Surfaces, in Geometry Processing for Design and Manufacturing (ed. R. E. Barnhill) (1992), SIAM, 45–86.
- [13] Hoschek, J. & Schneider, F.-J.: Approximate conversion and data compression of integral and rational B-spline surfaces, in Curves and Surfaces in Geometric Design, P.-J. Laurent, A. Le Méhauté, and L. L. Schumaker (eds.) (1994), A. K. Peters, 241–250.

- [14] Laurent-Gengoux, P. & Mekhilef, M.: Optimization of a NURBS representation, *Computer-Aided Design* Vol. 25 No. 11 (1993), 699–710.
- [15] Lyche, T. & Mørken, K.: Knot removal for B-spline curves and surfaces, *Computer Aided Geometric Design* 4 (1987), 217–230.
- [16] Ma, W. & Kruth, J.-P.: NURBS Curve and Surface Fitting and Interpolation, in *Mathematical Methods for Curves and Surfaces*, M. Dæhlen, T. Lyche, and L. L. Schumaker (eds.) (1995), Vanderbilt University Press, 305–314.
- [17] Ma, W. & Kruth, J.-P.: Parametrization of randomly measured points for least squares fitting of B-spline curves and surfaces, *Computer-Aided Design* Vol. 27 No. 6 (1995), 363–370.
- [18] Peternell, M. & Pottmann, H.: Designing Rational Surfaces with Rational Offsets, *Technical Report Nr. 28*, TU Wien (1995).
- [19] Piegl, L. & Tiller, W.: *The NURBS Book* (1995), Springer.
- [20] Pottmann, H. & Randrup, T.: Rotational and Helical Surface Approximation for Reverse Engineering, *Technical Report Nr. 43*, TU Wien (1997).
- [21] Pottmann, H., Peternell, M. & Ravani, B.: Contributions to computational line geometry, *Technical Report Nr. 50*, TU Wien (1998).
- [22] Pratt, M. J., Gault, R. J. & Ye, L.: On rational parametric curve approximation, *Computer Aided Geometric Design* 10 (1993), 363–377.
- [23] Qin, H. & Terzopoulos, D.: Dynamic NURBS swung surfaces for physics-based shape design, *Computer-Aided Design* Vol. 27 No. 2 (1995), 111–127.
- [24] Russo, R.: *Rationale Approximation von Trimmkurven*, Diplomarbeit TU Darmstadt (1996).
- [25] Schneider, F.-J.: *Interpolation, Approximation und Konvertierung mit rationalen B-Splines*, Dissertation TH Darmstadt (1992).
- [26] Schaal, H.: *Lineare Algebra und Analytische Geometrie* Bd. 2 (1976), Vieweg.

Lebenslauf

Persönliche Daten:	Bernhard Elsässer geb. 24. 6. 67 in Würzburg
Schul Ausbildung:	1973–1977 Grundschule in Rüsselsheim 1977–1979 Förderstufe in Groß-Gerau 1979–1986 Gymnasium mit Abitur in Groß-Gerau
Grundwehrdienst:	1986–1987 in Arolsen
Hochschulstudium:	1987–1994 Mathematik an der TU Darmstadt
wissenschaftliche Hilfskraft mit Abschluß:	seit 1995 an der TU Darmstadt